

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:04:54 ; Search time 11 Seconds  
(without alignments)  
27.566 Million cell updates/sec

Title: US-09-824-053-3

Perfect score: 77  
Sequence: 1 DLPSPRCVIASTNMF 16

Gaping table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PTCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCURS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
75	97.4	16	10	US-09-824-053-3	Sequence 3, Appl1
75	97.4	546	10	US-09-998-284-2	Sequence 2, Appl1
75	97.4	546	10	US-09-824-053-31	Sequence 31, Appl1
74	97.4	16	10	US-09-824-053-10	Sequence 10, Appl1
5	40	1884	10	US-09-785-770A-17	Sequence 17, Appl1
6	40	51.9	1907	US-09-785-770A-16	Sequence 16, Appl1
7	39	50.6	160	US-09-764-864-1020	Sequence 1020, Ap
8	39	50.6	442	US-09-731-872-286	Sequence 286, App
9	39	50.6	468	US-09-731-872-283	Sequence 283, App
10	39	50.6	468	US-09-933-561-2	Sequence 2, Appl1
11	38	49.4	393	US-10-090-185-10	Sequence 10, Appl1
12	38	49.4	609	US-09-815-242-10740	Sequence 10740, A
13	38	49.4	793	US-09-925-302-780	Sequence 780, App
14	38	49.4	4545	US-09-873-403-2	Sequence 2, Appl1
15	37	48.1	44	US-09-864-761-36590	Sequence 36590, A
16	37	48.1	119	US-09-867-550-246	Sequence 246, App
17	37	48.1	607	US-09-815-242-13379	Sequence 13379, A
18	37	48.1	607	US-09-815-242-13682	Sequence 13682, A
19	37	48.1	637	US-09-815-242-12058	Sequence 12058, A

20	36	46.8	388	9	US-09-975-139-9	Sequence 9, Appl1
21	36	46.8	669	10	US-09-801-196-28	Sequence 28, Appl
22	36	46.8	4636	10	US-09-835-996A-33	Sequence 33, Appl
23	35.5	46.1	516	9	US-10-067-534-3	Sequence 3, Appl1
24	35.5	46.1	520	9	US-10-067-534-4	Sequence 4, Appl1
25	35	45.5	91	10	US-09-864-761-36754	Sequence 36754, A
26	35	45.5	230	9	US-10-007-814-8	Sequence 8, Appl1
27	35	45.5	420	9	US-10-007-814-6	Sequence 6, Appl1
28	35	45.5	461	9	US-09-738-626-5041	Sequence 5041, Ap
29	35	45.5	503	9	US-10-007-814-2	Sequence 2, Appl1
30	35	45.5	504	9	US-10-007-814-4	Sequence 4, Appl1
31	34.5	44.8	419	10	US-09-741-669-415	Sequence 415, App
32	34	44.2	156	10	US-09-864-761-39604	Sequence 39604, A
33	34	44.2	217	9	US-09-738-626-5837	Sequence 5837, Ap
34	34	44.2	260	9	US-09-764-868-1030	Sequence 1030, Ap
35	34	44.2	451	9	US-09-738-626-3550	Sequence 3550, Ap
36	34	44.2	618	10	US-09-817-676A-14	Sequence 14, Appl
37	34	44.2	618	10	US-09-970-516-4	Sequence 4, Appl1
38	34	44.2	2273	10	US-09-995-542-12	Sequence 12, Appl
39	34	44.2	2310	10	US-09-995-542-10	Sequence 10, Appl
40	34	44.2	26926	9	US-09-759-508B-2	Sequence 2, Appl1
41	33	42.9	15	12	US-10-085-027-4	Sequence 4, Appl1
42	33	42.9	64	10	US-09-864-761-43603	Sequence 43603, A
43	33	42.9	64	10	US-09-764-871-1246	Sequence 1246, Ap
44	33	42.9	104	9	US-09-796-692-1793	Sequence 1793, Ap
45	33	42.9	104	9	US-09-796-692-2057	Sequence 2057, Ap

## ALIGNMENTS

RESULT 1  
US-09-824-053-3  
Sequence 3, Application US/09824053  
Patent No. US20020106725A1

## GENERAL INFORMATION:

APPLICANT: Peter Stougaard  
Ole Cal Hansen

TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A

METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hulton & Williams  
STREET: 1900 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006-1109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/824,053

FILING DATE: 03-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/669,304

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Stanislaus Aksman

REGISTRATION NUMBER: 28,562

REFERENCE/DOCKET NUMBER: <Unknown>

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 955-1926

TELEFAX: (202) 778-2201

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-824-053-3

Query Match 97.4%; Score 75; DB 10; Length 16;  
Best Local Similarity 100.0%; Pred. No. 5e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLPMSPRGVASNLKF 16  
|||||  
Db 1 DLPMSPRGVASNLKF 16

RESULT 2  
US-09-998-284-2  
Sequence 2, Application US/09998284  
Patent No. US20020106361A1  
GENERAL INFORMATION:  
APPLICANT: POULSEN, et al.  
TITLE OF INVENTION: COMPOSITION  
FILE REFERENCE: 674509-2035  
CURRENT APPLICATION NUMBER: US/09/998,284  
CURRENT FILING DATE: 2001-11-30  
PRIORITY APPLICATION NUMBER: PCT/IB00/00829  
PRIORITY FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: GB 9913050.2  
PRIOR FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 2  
LENGTH: 546  
TYPE: PRT  
ORGANISM: Chondrus crispus  
US-09-998-284-2

Query Match 97.4%; Score 75; DB 10; Length 546;  
Best Local Similarity 93.8%; Pred. No. 2.5e-05;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSPRGVASNLKF 16  
|||||  
Db 219 DLPMSPRGVASNLHF 234

RESULT 3  
US-09-824-053-31  
Sequence 31, Application US/09824053  
Patent No. US20020106725A1  
GENERAL INFORMATION:  
APPLICANT: Peter Stougaard  
Ole Cai Hansen  
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A  
METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hutton & Williams  
STREET: 1900 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006-1109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/824,053  
FILING DATE: 03-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/669,304

FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Stanislaus Aksman  
REGISTRATION NUMBER: 28,562  
REFERENCE/DOCKET NUMBER: <Unknown>  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 955-1926  
TELEFAX: (202) 778-2201  
TELEX: No. US20020106725A1e  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 31:  
US-09-824-053-31

Query Match 97.4%; Score 75; DB 10; Length 546;  
Best Local Similarity 93.8%; Pred. No. 2.5e-05;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSPRGVASNLKF 16  
|||||  
Db 219 DLPMSPRGVASNLHF 234

RESULT 4  
US-09-824-053-10  
Sequence 10, Application US/09824053  
Patent No. US20020106725A1  
GENERAL INFORMATION:  
APPLICANT: Peter Stougaard  
Ole Cai Hansen  
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A  
METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hutton & Williams  
STREET: 1900 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006-1109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/824,053  
FILING DATE: 03-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/669,304  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Stanislaus Aksman  
REGISTRATION NUMBER: 28,562  
REFERENCE/DOCKET NUMBER: <Unknown>  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 955-1926  
TELEFAX: (202) 778-2201  
TELEX: No. US20020106725A1e  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-824-053-10

Query Match 96.1%; Score 74; DB 10; Length 16;  
Best Local Similarity 93.8%; Pred. No. 7.4e-07;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSRGVYASNLXF 16  
|||||

Db 1 DLPMSRGVYASNLWF 16

RESULT 5  
US-09-785-770A-17  
; Sequence 17, Application US/09785770A  
; Patent No. US20020103360A1  
; GENERAL INFORMATION:  
; APPLICANT: Barnes, Thomas M.  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN  
; FILE REFERENCE: 07334-328001  
; CURRENT APPLICATION NUMBER: US/09/785,770A  
; CURRENT FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 09/387,462  
; PRIOR FILING DATE: 1999-09-01  
; PRIOR APPLICATION NUMBER: US 09/145,056  
; PRIOR FILING DATE: 1998-09-01  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 4.0  
; SEQ ID NO 17  
; LENGTH: 1884  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-785-770A-17

Query Match 51.9%; Score 40; DB 10; Length 1884;  
Best Local Similarity 43.8%; Pred. No. 90;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 DLPMSRGVYASNLXF 16  
|||||

Db 1804 DLPMSRGVYASNLWF 1819

RESULT 6  
US-09-785-770A-16  
; Sequence 16, Application US/09785770A  
; Patent No. US20020103360A1  
; GENERAL INFORMATION:  
; APPLICANT: Barnes, Thomas M.  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN  
; FILE REFERENCE: 07334-328001  
; CURRENT APPLICATION NUMBER: US/09/785,770A  
; CURRENT FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 09/387,462  
; PRIOR FILING DATE: 1999-09-01  
; PRIOR APPLICATION NUMBER: US 09/145,056  
; PRIOR FILING DATE: 1998-09-01  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 4.0  
; SEQ ID NO 16  
; LENGTH: 1907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-785-770A-16

Query Match 51.9%; Score 40; DB 10; Length 1907;  
Best Local Similarity 43.8%; Pred. No. 92;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 DLPMSRGVYASNLXF 16  
|||||

Db 1827 DLPMSRGVYASNLWF 1842

RESULT 7  
US-09-764-864-1020  
; Sequence 1020, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; PRIOR APPLICATION data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1020  
; LENGTH: 160  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-1020

Query Match 50.6%; Score 39; DB 10; Length 160;  
Best Local Similarity 70.0%; Pred. No. 8.8;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 DLPMSRGVYASNL 11  
:|:|

Db 137 DLPMSRGVYASNL 146

RESULT 8  
US-09-731-872-286  
; Sequence 286, Application US/09731872  
; Patent No. US20020102604A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Bouguetere, Lydie  
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
; FILE REFERENCE: 78.US3.REG  
; CURRENT APPLICATION NUMBER: US/09/731,872  
; CURRENT FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/169,629  
; PRIOR FILING DATE: 1999-12-08  
; PRIOR APPLICATION NUMBER: US 60/187,470  
; PRIOR FILING DATE: 2000-03-06  
; NUMBER OF SEQ ID NOS: 482  
; SOFTWARE: PatentIn Ver. 4.0  
; SEQ ID NO 286  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -21..-1  
; NAME/KEY: UNSURE  
; LOCATION: 132  
; OTHER INFORMATION: Xaa = Pro,Arg  
US-09-731-872-286

Query Match 50.6%; Score 39; DB 10; Length 442;  
Best Local Similarity 37.5%; Pred. No. 27;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 DLPMSRGVYASNLXF 16  
:|:|

Db 176 DLPMSRGVYASNLWF 191

RESULT 9  
US-09-731-872-283  
; Sequence 283, Application US/09731872

Patent No. US20020102604A1  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Joberet, Severin  
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
FILE REFERENCE: 78.053.REG  
CURRENT APPLICATION NUMBER: US/09/731,872  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,629  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: US 60/187,470  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 482  
SOFTWARE: Patent.pm  
SEQ ID NO 283  
LENGTH: 468  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -21...-1  
9-731-872-283

Query Match 50.6%; Score 39; DB 10; Length 468;  
Best Local Similarity 37.5%; Pred. No. 29;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 DLPMSRGVIAINLXF 16  
DB 176 NLPLSPQGTVRTAVEF 191

RESULT 10  
US-09-933-561-2  
Sequence 2, Application US/09933561  
Patent No. US2002010664A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti G.  
APPLICANT: Kaser, Matthew R.  
APPLICANT: Baugh, Mariah R.  
TITLE OF INVENTION: TAPASIN-LIKE PROTEIN  
FILE REFERENCE: PC-0002-1 CIP  
CURRENT APPLICATION NUMBER: US/09/933,561  
CURRENT FILING DATE: 2001-08-20  
PRIOR APPLICATION NUMBER: 09/292,097  
PRIOR FILING DATE: 1999-04-14  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 468  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: 103348CD1  
09-933-561-2

Query Match 50.6%; Score 39; DB 10; Length 468;  
Best Local Similarity 37.5%; Pred. No. 29;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 DLPMSRGVIAINLXF 16  
DB 176 NLPLSPQGTVRTAVEF 191

RESULT 11  
US-10-090-185-10  
Sequence 10, Application US/10090185  
Publication No. US20020197647A1  
GENERAL INFORMATION:  
APPLICANT: Zhang, Xiaokui  
APPLICANT: Wieszczynska, Melissa H

APPLICANT: Horvath, Curt M  
APPLICANT: Darnell Jr., James E  
TITLE OF INVENTION: METHODS FOR IDENTIFYING MODULATORS OF TRANSCRIPTIONAL ACTIVITY  
TITLE OF INVENTION: INTERACTIONS  
FILE REFERENCE: 600-1-253  
CURRENT APPLICATION NUMBER: US/10/090,185  
CURRENT FILING DATE: 2002-03-04  
PRIOR APPLICATION NUMBER: 09/387,418  
PRIOR FILING DATE: 1999-08-31  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-090-185-10

Query Match 49.4%; Score 38; DB 9; Length 393;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLPMSPR 7  
DB 346 DLPMSPR 352

RESULT 12  
US-09-815-242-10740  
Sequence 10740, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes In  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10740  
LENGTH: 609  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(609)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-10740

Query Match 49.4%; Score 38; DB 10; Length 609;  
Best Local Similarity 43.8%; Pred. No. 57;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 DLPMSRGVIASTNMF 16  
Db 432 DIPADRGVQJEVSF 447

RESULT 13  
US-09-925-302-780  
; Sequence 780, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 780  
; LENGTH: 793  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-302-780

Query Match 49.4%; Score 38; DB 10; Length 793;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLPMSR 7  
Db 746 DLPMSR 752

RESULT 14  
US-09-873-403-2  
; Sequence 2, Application US/09873403  
; Patent No. US20020028207A1  
; GENERAL INFORMATION:  
; APPLICANT: Strivastava, Pramod K  
; TITLE OF INVENTION: COMPLEXES OF ALPHA (2) MACROGLOBULIN AND ANTIGENIC  
; FILE REFERENCE: 8449-178  
; CURRENT APPLICATION NUMBER: US/09/873,403  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/625,139  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/209,266  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 4545  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-873-403-2

Query Match 49.4%; Score 38; DB 10; Length 4545;  
Best Local Similarity 54.3%; Pred. No. 5,2e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 DLPMSRGVIA 11  
Db 2460 DIPQPMGITA 2470

RESULT 15  
US-09-864-761-36590  
; Sequence 36590, Application US/09864761  
; Patent No. US20020048763A1

; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36590  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens

; OTHER INFORMATION: MAP TO AC006367.3  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2  
; OTHER INFORMATION: SWISSPROT HIT: P77672, EVALUATE 2.10e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: BE960561.1, EVALUATE 1.80e+00  
US-09-864-761-36590

Query Match 48.1%; Score 37; DB 10; Length 44;  
Best Local Similarity 58.3%; Pred. No. 4.6;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PMSRGVIASTNL 14

Db 4 PMPRGVAVSL 15

## RESULT 16

US-09-867-550-246  
; Sequence 246, Application US/09867550  
; Patent No. US20020082206A1

## GENERAL INFORMATION:

APPLICANT: Leach, Martin D.  
APPLICANT: Mehriban, Foad,  
APPLICANT: Conley, Pamela  
APPLICANT: Law, Debbie

APPLICANT: Topper, James

TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and

FILE REFERENCE: 21402-013 (Curra-313)

CURRENT APPLICATION NUMBER: US/09/867,550

PRIOR FILING DATE: 2001-09-20

PRIOR APPLICATION NUMBER: US2002008,427

PRIOR FILING DATE: 2000-05-30

NUMBER OF SEQ ID NOS: 2125

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 246

LENGTH: 119

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)

OTHER INFORMATION: wherein Xaa may be any one of Arg or Gln or Leu or Pro or Ser or

OTHER INFORMATION: Thr or Lys or Ala or Val or Glu or Gly

US-09-867-550-246

Query Match 48.1%; Score 37; DB 10; Length 119;  
Best Local Similarity 45.5%; Pred. No. 14;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Y 4 MSPRGVAVSL 14  
1:|||||:  
Db 51 MAPRGVAVSV 61

## RESULT 17

US-09-815-242-13379

; Sequence 13379, Application US/09815242

; Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13379  
; LENGTH: 607  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13379

Query Match 48.1%; Score 37; DB 10; Length 607;  
Best Local Similarity 37.5%; Pred. No. 84;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Y 1 DLPMPRGVAVSLX 16  
1:|||||:  
Db 431 DIPMPRGVAVSLX 446

## RESULT 18

US-09-815-242-13682

; Sequence 13682, Application US/09815242

; Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13682

LENGTH: 607

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-815-242-13682

Query Match 48.1%; Score 37; DB 10; Length 607;  
Best Local Similarity 37.5%; Pred. No. 84;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Y 1 DLPMPRGVAVSLX 16  
1:|||||:  
Db 431 DIPMPRGVAVSLX 446

## RESULT 19

US-09-815-242-12058

; Sequence 12058, Application US/09815242

; Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 12038  
 LENGTH: 637  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-815-242-12058

Query Match 48.18; Score 37; DB 10; Length 637;  
 Best Local Similarity 43.88; Pred. No. 89;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 DLPMPRGVIAASNLXF 16  
 Db 461 DIPAPRGVPOIEYTF 476

RESULT 20  
 US-09-975-139-9  
 Sequence 9, Application US/09975139  
 Patent No. US20020155460A1  
 GENERAL INFORMATION:  
 APPLICANT: Genencor International, Inc.  
 APPLICANT: Schellenberger, Volker  
 APPLICANT: Naki, Donald  
 APPLICANT: Morrison, Thomas B.  
 TITLE OF INVENTION: INFORMATION RICH LIBRARIES  
 FILE REFERENCE: 23623-7060  
 CURRENT APPLICATION NUMBER: US/09/975,139  
 CURRENT FILING DATE: 2001-10-10  
 PRIOR APPLICATION NUMBER: US 60/239,476  
 PRIOR FILING DATE: 2000-10-10  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9  
 LENGTH: 388  
 TYPE: PRT  
 ORGANISM: Yersinia enterocolitica  
 FEATURE:  
 OTHER INFORMATION: AmpC protein  
 US-09-975-139-9

Query Match 46.88; Score 36; DB 9; Length 388;  
 Best Local Similarity 54.58; Pred. No. 76;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 PMSPRGVIAASN 13  
 Db 304 PINPQGVIAADS 314

Search completed: January 2, 2003, 12:10:23  
 Job time : 12 secs



1  
2  
3  
4





DR WPI; 1997-052332/05.  
XX Recombinant production of polypeptide having hexose oxidase activity  
PT - used in food preparations as antibacterial and antioxidant agent  
XX  
PS Claim 6; Page 99; 124pp; English.  
XX  
CC AAM20068-75 are small peptides derived from hexose oxidase (HO) of the  
CC marine algae species Chondrus crispus. Hexose oxidase and peptides  
CC having HO activity are useful in the production of food products, e.g.  
CC dairy products, starch-containing food products (dough) and non-dairy  
CC beverages. HO and active peptide fragments have antimicrobial and  
CC antioxidant properties and act by removing all the oxygen in a food  
CC packaging. HO and peptides of HO can also be used in an animal feed,  
CC especially silage. Further uses are to reduce, or analyse, the sugar  
CC content in a food, in the production of cosmetics, tooth care products  
CC or a pharmaceutical product and in lactone production. HO can be  
CC recombinantly produced in industrially appropriate quantities, and at  
CC a quality and purity level which renders polypeptide suitable for  
CC industrial purposes.

Sequence 16 AA;  
Query Match 97.4%; Score 75; DB 18; Length 16;  
Best Local Similarity 100.0%; Pred. No. 9.9e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 DLPMSPRGVASNLXF 16  
Db 1 DLPMSPRGVASNLXF 16

RESULT 2  
AAM20076  
ID AAM20076 standard; Protein: 546 AA.  
XX  
AC AAM20076;  
XX  
DT 12-SEP-1997 (first entry)  
XX  
DE Hexose oxidase, an antimicrobial/antioxidant agent.  
XX  
XX Hexose oxidase; Chondrus crispus; marine algae; recombinant;  
KW antimicrobial; antioxidant; food preparation; dairy product; starch;  
KW beverage; animal feed; silage; sugar reduction; cosmetics; dental;  
KW toothpaste; dough; lactone production.  
XX  
OS Chondrus crispus.  
XX  
PN WO9640935-A1.  
XX  
XX 19-DEC-1996.  
XX  
XX 04-JUN-1996; 96WO-DK00238.  
XX  
XX 07-JUN-1995; 95US-0476910.  
XX  
XX (BIOT-) BIOTEKNOLOGISK INST.  
XX  
XX Hansen OC, Stougaard P;  
XX  
XX WPI; 1997-052332/05.  
XX  
XX N-PSDB; AAT76552.  
XX  
PT Recombinant production of polypeptide having hexose oxidase activity  
PT - used in food preparations as antibacterial and antioxidant agent  
XX  
XX Example 3.5; Page 95-97; 124pp; English.  
XX  
XX AAM20076 shows the hexose oxidase (HO) enzyme of the marine algae  
CC species Chondrus crispus. HO is useful in the production of food  
CC products, e.g. dairy products, starch-containing food products (dough)  
CC and non-dairy beverages. HO and active peptide fragments have

CC antimicrobial and antioxidant properties and act by removing all the  
CC oxygen in a food packaging. HO and peptides of HO can also be used in an  
CC animal feed, especially silage. Further uses are to reduce, or analyse,  
CC the sugar content in a food, in the production of cosmetics, tooth care  
CC products or a pharmaceutical product and in lactone production. HO can  
CC be recombinantly produced in industrially appropriate quantities, and at  
CC a quality and purity level which renders the polypeptide suitable for  
CC industrial purposes.

Sequence 546 AA;  
Query Match 97.4%; Score 75; DB 18; Length 546;  
Best Local Similarity 93.8%; Pred. No. 5.3e-05;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 DLPMSPRGVASNLXF 16  
Db 219 DLPMSPRGVASNLHF 234

RESULT 3  
AAY83619  
ID AAY83619 standard; Protein: 546 AA.  
XX  
AC AAY83619;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Synthetic hexose oxidase.  
XX  
XX Hexose oxidase; production; fermentation; synthetic; modification;  
KW prokaryote; eukaryote.  
XX  
OS Synthetic.  
XX  
XX EP1008651-A2.  
XX  
XX 14-JUN-2000.  
XX  
XX 01-DEC-1999; 99EP-0204068.  
XX  
XX 09-DEC-1998; 98DK-0001630.  
XX  
XX (BIOT-) BIOTEKNOLOGISK INST.  
XX  
XX Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;  
XX  
XX WPI; 2000-389309/34.  
XX  
XX N-PSDB; AAZ94011.  
XX  
XX Nucleic acid fragment useful for producing large amounts of hexose  
PT oxidase comprises nucleotide sequence encoding hexose oxidase that is  
PT modified by at least one codon  
XX  
XX Disclosure; Page 28-29; 42pp; English.  
XX  
XX Wild type hexose oxidase can be produced by isolating a nucleotide  
CC sequence coding for hexose oxidase naturally produced by an organism  
CC and modifying the sequence so that the modified sequence is  
CC expressed, under identical conditions, at a level at least 10% higher  
CC than the non-modified sequence. The method is useful for producing  
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods  
CC of producing hexose oxidase were to isolate the enzyme from a source  
CC that naturally produces the enzyme, using host organisms which  
CC produce a relatively low level of expression. Therefore, industrial  
CC production of the enzyme is not feasible using this method. This is a  
CC much improved method which produces at least 250 mg enzyme per liter  
CC of fermentation medium.  
XX  
XX Sequence 546 AA;  
XX  
XX Query Match 97.4%; Score 75; DB 21; Length 546;  
XX  
XX Best Local Similarity 93.8%; Pred. No. 5.3e-05;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DLPMSRGTASNLXF 16  
 |||||||  
 DB 219 DLPMSRGTASNLHF 234

## RESULT 4

AAU02192  
 ID AAU02192 standard; Protein; 546 AA.

AC AAU02192;

DT 12-SEP-2001 (first entry)

Synthetic hexose oxidase (HOX) amino acid sequence.

HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;  
 food manufacturing; beverage; detergent; baking; dough improving agent;  
 D-hexose:O2-oxidoreductase; EC 1.1.3.5.

OS Chondrus crispus.  
 OS Synthetic.

PN WO200138544-A1.

PD 31-MAY-2001.

PF 24-NOV-2000; 2000WO-IB01886.

PR 24-NOV-1999; 99GB-0027801.

PA (DANI-) DANISCO AS.

PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;

PI zargahl MR;

DR WPI: 2001-367695/38.

DR N-PSDB: MAS06173.

PT Releasing soluble or membrane associated intracellular protein from a  
 cell for manufacturing food, comprises contacting the cell with a  
 membrane extracting composition and causing the protein to be released  
 in soluble form -

PS Claim 16; Fig 6; 108pp; English.

The sequence represents the amino acid sequence of synthetic hexose  
 oxidase (D-hexose:O2-oxidoreductase, EC 1.1.3.5), also referred to as HOX.  
 The native HOX gene was altered using site-directed mutagenesis in order  
 to match the codon usage to known codon preferences of biotechnologically  
 relevant yeasts, such as *Pichia sp.*, to facilitate high level production  
 of HOX in these organisms. The invention involves a method for releasing  
 a soluble or membrane associated intracellular protein of interest (POI)  
 from a cell involving contacting a cell comprising a soluble or membrane  
 associated intracellular POI with a membrane extracting composition (I)  
 and causing the POI to be released from the cell in a soluble form. The  
 method is useful for releasing POI, such as an interleukin I receptor  
 antagonist (IL-1ra) which involves contacting a transformed cell  
 comprising IL-1ra with (I) and causing IL-1ra to be released from the  
 transformed cell, in a soluble form. The method is also useful for  
 screening mutated cells or transformed cells producing elevated levels of  
 intracellular POI. The method is used to release a POI for manufacturing  
 food products, such as beverages, preparation of detergents, and in  
 baking as a dough improving agent. The method obtains a fast, specific  
 and economically efficient extraction of a soluble or membrane associated  
 intracellular POI without the use of conventional cell disruption  
 techniques. The resulting cell extract contains less contaminating  
 intracellular DNA and is relatively free of cell wall fragments. The  
 intracellular POI can be recovered from a eukaryotic host organism such  
 as yeast, before glycosylation takes place. The method can be used to  
 prevent contact of intracellular POI with the extracellular growth  
 medium.

XX Sequence 546 AA;

SO Query Match 97.4%; Score 75; DB 22; Length 546;  
 Best Local Similarity 93.8%; Pred. No. 5.3e-05;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DLPMSRGTASNLXF 16  
 |||||||  
 DB 219 DLPMSRGTASNLHF 234

## RESULT 5

AAU02192  
 ID AAU02192 standard; Protein; 546 AA.

AC AAU02192;

DT 23-MAR-2001 (first entry)

DE Chondrus crispus hexose oxidase enzyme protein.

KM Hexose oxidase; marine alga; anti-fouling.

OS Chondrus crispus.

PN WO200075293-A2.

PD 14-DEC-2000.

PF 02-JUN-2000; 2000WO-IB00829.

PR 04-JUN-1999; 99GB-0013050.

PA (DANI-) DANISCO AS.

PI Poulsen CH, Kragh KM;

DR WPI: 2001-112148/12.

PT New anti-fouling composition, useful as a coating for treating  
 different surfaces, e.g. outdoor woodwork, external surface of a  
 central heating system, or a hull of a marine vessel -

PS Claim 5; Page 35-36; 36pp; English.

The present invention relates to a new anti-fouling composition. The  
 composition involves a surface coating material, a hexose oxidase  
 enzyme obtained from a marine organism and a substrate for the  
 enzyme. The anti-fouling composition is useful as a coating formulated  
 for treating a surface, e.g. outdoor wood work, external surface of a  
 central heating system, or a hull of a marine vessel. It is  
 also useful as an anti-fouling agent for marine structures exposed to  
 seawater flora and fauna.

SO Sequence 546 AA;

Query Match 97.4%; Score 75; DB 22; Length 546;  
 Best Local Similarity 93.8%; Pred. No. 5.3e-05;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DLPMSRGTASNLXF 16  
 |||||||  
 DB 219 DLPMSRGTASNLHF 234

## RESULT 6

AAU02192  
 ID AAU02192 standard; Protein; 187 AA.

AC AAU02192;

DT 02-JUL-2002 (first entry)

XX	Streptococcus polypeptide SEQ ID NO 10084.
DE	
XX	Streptococcus: GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KX	group A Streptococcus; Streptococcus pyogenes; antibacterial;
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.
KX	
OS	Streptococcus agalactiae.
XX	
PN	WO200234771-A2.
PD	
XX	02-MAY-2002.
XX	
PE	29-OCT-2001; 2001WO-GB04789.
XX	
PR	27-OCT-2000; 2000GB-0026333.
PR	24-NOV-2000; 2000GB-0028727.
ER	07-MAR-2001; 2001GB-0005640.
XX	
PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	
XX	Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX	Tettelin H;
DR	WPI; 2002-352536/38.
DR	N-PSDB; ABN71085.
XX	
PT	New Streptococcus protein for the treatment or prevention of infection
PT	or disease caused by Streptococcus bacteria, such as meningitis, and
PT	for detecting a compound that binds to the protein -
XX	
PS	Claim 1; Page 4137; 4535pp; English.
XX	
XX	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1) given in
CC	the specification. The proteins have antibacterial and antiinflammatory
CC	activities. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins.
XX	
XX	Sequence 187 AA:
XX	
XX	Every Match 57.1%; Score 44; DB 23; Length 187;
XX	Best Local Similarity 75.0%; Pred. No. 6.1;
XX	Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY	5 SPRGVYASNLXF 16
DB	13 SPRGVYANKLVF 24
XX	
XX	RESULT 7
XX	AAV91319
XX	AAV91319 standard; Protein; 230 AA.
XX	AAV91319;
XX	30-MAY-2000 (first entry)
XX	
DE	Group B Streptococcus protein sequence SEQ ID NO:47.
XX	

xx	OS	Streptococcus agalactiae.
xx	PN	WO200006736-A2.
xx	PD	10-FEB-2000.
xx	PF	27-JUL-1999; 99WO-GB02444.
xx	PR	27-JUL-1998; 98GB-0016335.
xx	PR	19-MAR-1999; 99US-0125163.
xx	PA	(MICR-) MICROBIAL TECHNICS LTD.
xx	PI	Le Page RWF, Wells JM, Hanniffy SB;
xx	DR	WPI; 2000-195299/17.
xx	PT	New Group B Streptococcus protein, useful as vaccine, for diagnosis of
xx	PT	Streptococcal infections and for screening of antibodies or affibodies
xx	PS	Claim 1; Fig 1; 123pp; English.
xx	CC	AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
xx	CC	in AA911275 to AA911343) isolated from Group B Streptococcus (GBS), also
xx	CC	known as Streptococcus agalactiae. The GBS polynucleotides and
xx	CC	polypeptides have antibacterial activity. Immunogenic compositions
xx	CC	comprising GBS polynucleotides or polypeptides can be used as vaccines
xx	CC	and for the treatment or prophylaxis of GBS infection. The
xx	CC	polynucleotides and polypeptides can also be used in the detection of GBS
xx	CC	and for screening DNA encoding bacterial cell envelope associated or
xx	CC	secreted antigens in gram positive bacteria. AAA05873 to AAA05941
xx	CC	represent primers used in the exemplification of the present invention.
xx	SEQ	Sequence 230 AA;
xx	Query Match	57.1%; Score 44; DB 21; Length 230;
xx	Best Local Similarity	75.0%; Pred. No. 7.7;
xx	Matches 9; Conservative	1; Mismatches 2; Indels 0; Gaps 0;
OY	5 SPRGVASLTXF 16	
	:	
DB	56 SPRGVIANKLVE 67	
RESULT 8		
ABP28247		
ID	ABP28247 standard; Protein; 230 AA.	
xx	AC	ABP28247;
xx	DT	02-JUL-2002 (first entry)
xx	DE	Streptococcus polypeptide SEQ ID NO 5670.
xx	KW	Streptococcus; GAS; GBS; group B Streptococcus; Streptococcus agalactiae;
xx	KW	group A streptococcus; Streptococcus pyogenes; antibacterial;
xx	KW	ant inflammatory; infection; vaccine; meningitis; gene therapy.
xx	OS	Streptococcus agalactiae.
xx	PN	WO200234771-A2.
xx	PD	02-MAY-2002.
xx	PF	29-OCT-2001; 2001WO-GB04789.
xx	PR	27-OCT-2000; 2000GB-0026333.
xx	PR	24-NOV-2000; 2000GB-0028727.
PR	07-MAR-2001; 2001GB-0005640.	

XX (CHTR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX WPI: 2002-352536/38.  
 DR N-PSDB: ABN68878.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 Claim 1; Page 3729; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus (Streptococcus agalactiae) or group A streptococcus (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

SO Sequence 230 AA;  
 Query Match 57.1%; Score 44; DB 23; Length 230;  
 Best Local Similarity 75.0%; Pred. No. 7.7;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Oy 5 SPRGVIANSLXK 16  
 |||||: ||  
 Db 56 SPRGVIANSLXK 67

RESULT 9  
 ABG07191  
 ABG07191 standard; Protein: 2506 AA.

XX ABG07191:  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #7182.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS71378.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID NO 37550; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at fip.wipo.int/pub/published\_pcl\_sequences.

SO Sequence 2506 AA;  
 Query Match 54.5%; Score 42; DB 22; Length 2506;  
 Best Local Similarity 53.3%; Pred. No. 2.6e+02;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 Oy 2 LPMSPRGVIANSLXK 16  
 |||||: ||  
 Db 745 LPAPRGVQRLRLK 759

RESULT 10  
 ABG30064  
 ID ABG30064 standard; Protein: 2506 AA.  
 XX

XX ABG30064:  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #30055.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS94251.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 60423; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (I) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

**SQ**      **Sequence**      **2506 AA;**

Query Match	54.58;	Score 42;	DB 22;	Length 2506;
Best Local Similarity	53.38;	Pred. No. 2.6e+02;		
Matches	8;	Conservative	6;	Indels 0; Gaps 0

QY 2 LPMSPRGVIASNLXF 16

Db 745 LPAPRGVQRRLF 759

## RESULT 11

ID ABG23334 standard; Protein; 185 AA.

... AC ABG23334 ;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #233325.

**kw** Human; chromosome mapping; gene mapping; gene therapy; forensic food supplement; medical imaging; diagnostic; genetic disorder.

**Homo sapiens**

PN WO200175067-A2

11-OCT-2001  
PD

30-MAR-2001; 2001WO-US08631.

AA 31-MAR-2000; 2000US-0540217  
PR

XX 23-AUG-2000, 200005-064510/  
XX

XX PA (HISE-) HISEQ INC.

P1 Drmanac R1, Liu C, Tang Y1  
XX

DR WPI; 2001-639362/73  
DR N-PSDB: AAC87521

**XX**

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 53693; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (I) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence 185 AA

Query Match	51.98;	Score 40;	DB 22;	Length 185;
Best Local Similarity	70.08;	Pred. No. 32;		
Matches 7; Conservative	2;	Mismatches	1;	Indels 0; Gaps 0;

QY 3 PMSPRGV IAS 12

Db 50 PMSPSGVSS 59

## RESULT 12

ID ABP41781 standard; Protein; 245 AA

AC ABP41781

DT 22-AUG-2002 (first entry)

Human ovarian antigen HLDAS11, SEQ ID NO:2913.

KW Human ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory system disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive.

aa Homo sapiens

PN WO200200677-A1

03-JAN-2002  
PD  
AA

AA  
PE 07-JUN-2001: 2001WO-US18569

XX  
PR 07-JUN-2000: 2000JTS-209467PXX  
DA (HIMA - ) HITMAN GENOME SCT INC

XX  
F  
C  
P  
C

XX WPI: 2002-147878/19.  
DR N-PSDB: ABQ54858.

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
PT ovarian cancer), immune disorders, cardiovascular disorders and  
PT neurological diseases -

PS Claim 11; SEQ ID No 2913; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovarian and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents a human ovarian antigen of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

XX Sequence 245 AA:

Query Match 51.9%; Score 40; DB 23; Length 245;  
Best Local Similarity 43.8%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

1 DLPSPRGVIASNLX 16  
||| ||| : :  
165 DLPHPRGFLPGHAP 180

RESULT 13

AAU72763 standard; Protein: 255 AA.

AAU72763:

26-FEB-2002 (first entry)

Tomato DMT protein related sequence #2.

XX Demeter: DMT; Atropos: ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.

XX Lycopersicon esculentum.

XX MO200180626-A1.

XX 01-NOV-2001.

XX 23-APR-2001; 2001MO-US13059.

XX 21-APR-2000; 2000US-0553690.

XX (REGC ) UNIV CALIFORNIA.

XX Fischer RL, Choi Y, Hannon M, Okumuro JK, Tatartanova TV;

XX WPI: 2002-055307/07.

XX N-PSDB: AAS96713.

PT New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C -

PS Disclosure: Page 96; 109pp; English.

XX The invention relates to an isolated polynucleotide sequence or their  
CC complement encoding a polypeptide having a sequence at least 40%  
CC identical to DMT (DEMETER, previously known as ATRPOS (ATR) Domain A, B  
CC or C or their combinations. Also included are an expression  
CC cassette comprising the polynucleotide or comprising a heterologous  
CC polynucleotide under the control of a promoter at least 70% identical to  
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
CC region of DMT, a host cell comprising an exogenous polynucleotide  
CC encoding a DMT-like protein and a transgenic plant comprising a  
CC polynucleotide encoding a DMT-like protein. The expression cassette is  
CC useful for modulating transcription. The method comprises introducing  
CC the cassette into a host cell preferably Agrobacterium by sexual  
CC cross, and selecting a host cell with modulated transcription, where  
CC the protein is capable of exhibiting at least one of the following  
CC biological activities, which include enhanced expression of the  
CC protein in a plant results in a delay in flowering time, introduction of  
CC the protein into a cell results in modulation of methylation of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expressing of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX

XX Sequence 255 AA:

Query Match 51.9%; Score 40; DB 23; Length 255;  
Best Local Similarity 50.0%; Pred. No. 46;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 DLPSPRGVIASNL 14  
||| ||| : :  
DB 120 DLPSPSSSVSSST 133

RESULT 14

AAAM25602 standard; Protein: 1193 AA.

AAAM25602:

16-OCT-2001 (first entry)

Human protein sequence SEQ ID NO:1117.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW anti-inflammatory; antirheumatic; antiallergic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; viticidic;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; anemic; anemia;  
KW antiagregant; haemostatic; vulnary; antileuk; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cystostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;

KM allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KM neurological disorder.  
 OS Homo sapiens.  
 XX MO200153455-A2.  
 XX  
 XX 26-JUL-2001.  
 XX  
 XX 22-DEC-2000; 2000MO-US35017.  
 XX  
 XX 23-DEC-1999; 99US-0471275.  
 XX 21-JAN-2000; 2000US-048725.  
 XX 25-APR-2000; 2000US-0552317.  
 XX  
 XX (HYSEQ-) HYSEQ INC.  
 XX Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI: 2001-457603/49.  
 XX N-PSDB; AAH9543.  
 PT Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 PS Claim 20; Page 230; 1217pp; English.  
 XX  
 XX AAH95166 to AAH99904 encode the human proteins given in AAH25225 to  
 CC AAH25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antineumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antidiabetic; haemostatic; vulnary;  
 CC anticancer; osteopathic; dermatological; antiallergic; antisthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.  
 CC  
 XX  
 XX Sequence 1193 AA;  
 SQ  
 PT Query Match 51.9%; Score 40; DB 22; Length 1193;  
 PT Best Local Similarity 43.8%; Pred. No. 2.6e+02;  
 PT Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 DLPMSPRGVIASNLXF 16  
 DB 1113 DLPMPRGFLPGHAPF 1128  
 |||: |||: : |  
 RESULT 15  
 ABG61824  
 ID ABG61824 standard; Protein: 1193 AA.  
 XX  
 XX ABG61824;  
 AC  
 XX 15-AUG-2002 (first entry)  
 XX  
 XX Prostate cancer-associated protein #25.  
 DE  
 XX Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.  
 KM  
 XX

OS Mammalia.  
 XX  
 XX WO200230268-A2.  
 PN  
 XX 18-APR-2002.  
 XX  
 XX PD  
 XX 12-OCT-2001; 2001MO-US32045.  
 XX  
 XX 13-OCT-2000; 2000US-0687576.  
 XX 08-DEC-2000; 2000US-0733288.  
 XX 08-DEC-2000; 2000US-0733742.  
 XX 24-JAN-2001; 2001US-263957P.  
 XX 16-MAR-2001; 2001US-276791P.  
 XX 16-MAR-2001; 2001US-276888P.  
 XX 06-APR-2001; 2001US-281922P.  
 XX 24-APR-2001; 2001US-286214P.  
 XX 30-APR-2001; 2001US-0847046.  
 XX 04-MAY-2001; 2001US-288589P.  
 XX  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 XX Gish KC, Mack DH, Wilson KE, Afar D, Hevez P;  
 XX  
 XX WPI: 2002-471335/50.  
 XX N-PSDB; ABK92139.  
 DR  
 XX  
 XX Detecting a prostate cancer-associated transcript in a cell in a  
 PT patient, useful for diagnosing prostate cancer (PC) or screening  
 PT modulators of PC, by determining if prostate cancer-associated genes  
 PT are expressed in a prostate tissue -  
 PS Claim 27; Page 321; 436pp; English.  
 XX  
 XX The present invention relates to methods of detecting a prostate  
 CC cancer-associated transcript in a cell from a patient. The method  
 CC comprises contacting a biological sample from the patient with  
 CC prostate cancer-associated polynucleotides (designated PC genes) that  
 CC selectively hybridise to a sequence that is at least 80% identical  
 CC to them. The prostate cancer-associated polynucleotide sequences  
 CC are differentially expressed in prostate tumour tissue or in  
 CC prostate cancer and are derived from the tissues of various  
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
 CC The methods of the invention are useful for diagnosing and treating  
 CC prostate cancer in mammals. The prostate cancer-associated genes are  
 CC useful for diagnosing or treating prostate cancer, as well as for  
 CC identifying modulators of prostate cancer or agents that inhibit  
 CC prostate cancer. The nucleic acid sequences are particularly useful  
 CC in gene therapy as a vaccine or in antisense applications.  
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins.  
 CC  
 XX  
 XX Sequence 1193 AA;  
 SQ  
 PT Query Match 51.9%; Score 40; DB 23; Length 1193;  
 PT Best Local Similarity 43.8%; Pred. No. 2.6e+02;  
 PT Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 DLPMSPRGVIASNLXF 16  
 DB 1113 DLPMPRGFLPGHAPF 1128  
 |||: |||: : |  
 RESULT 16  
 AAU32407  
 ID AAU32407 standard; Protein: 1194 AA.  
 XX  
 XX AAU32407;  
 AC  
 XX 18-DEC-2001 (first entry)  
 XX  
 XX Novel human secreted protein #2898.  
 DE  
 XX Human; vaccination; gene therapy; nutritional supplement;  
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 XX



us-09-824-053-3.rag

XX	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmamac RT, Liu C, Tang YT;
DR	WPI; 2001-639362/73.
XX	N-PSDB: AAS88229.
PT	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PS	Claim 20; SEQ ID No 54401; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or as
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
XX	at <a href="http://wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a> .
XX	
SO	Sequence 70 AA:
	Query Match 50.6%; Score 39; DB 22; Length 70;
	Best local Similarity 70.0%; Pred. No. 16;
	Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0.
QY	2 LPMSPRGVTA 11
	:1:11 1111
DB	42 IPLSPAGVIA 51
RESULT 18	
AAU22735	
ID	AAU22735 standard; Protein; 84 AA.
AC	AAU22735;
XX	
XX	17-DEC-2001 (first entry)
DE	Human prostate cancer antigen, Seq ID No 254.
OS	Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;
KW	reproductive system; chromosomal marker; forensics; urinary disorder;
KM	chronic nephritis; blood-related disorder; thrombosis.
OS	Homo sapiens.
XX	
PN	WO200155316-A2.
XX	
PD	02-AUG-2001.
XX	
XX	17-JAN-2001; 2001WO-US01328.
XX	



PT and also for testing and detection e.g. diagnosis -

XX Claim 11; SEQ ID No 254; 546bp; English.

XX The invention relates to novel isolated human prostate cancer antigen

CC polynucleotides (I) and polypeptides (II); (I) and (II) are useful for

CC preventing, treating or ameliorating a medical condition when

CC administered; (I), (II) and the antibody to (II) are useful for treating,

CC preventing and/or prognosing disorders related to the reproductive

CC system including prostate cancers; urinary disorders e.g. chronic

CC nephritis; and blood-related disorders e.g. thrombosis; (II) can be used

CC for testing and detection e.g. as a chromosomal marker and in forensics.

CC (I) and the anti-(II) antibody can be used in testing and detection in

CC immunoassays. AAU22702-AAU22913 represent the human prostate cancer

CC antigen amino acid sequences, and related amino acid sequences of the

CC invention. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at: ftp.wipo.int/pub/published\_pct\_sequences.

XX

SO Sequence 84 AA:

Query Match 50.68; Score 39; DB 22; Length 84;

Best Local Similarity 70.08; Pred. No. 20;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LPMSPRCVIA 11

DB 52 IPLSPAGVIA 61

RESULT 19

AAU94707

ID AAU94707 standard; Protein: 84 AA.

XX AAU94707:

XX 21-NOV-2001 (first entry)

DT Human reproductive system related antigen SEQ ID NO: 3365.

DE Human reproductive system related antigen; reproductive system disorder;

KW cancer; gene therapy.

XX Homo sapiens.

OS

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WC-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226581.

PR 22-AUG-2000; 2000US-0226688.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 12-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 21-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 25-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239335.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.



PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0237340
PR	13-OCT-2000	2000US-0239350
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240961
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	17-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249246
PR	17-NOV-2000	2000US-0249247
PR	17-NOV-2000	2000US-0249249
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250150
PR	01-DEC-2000	2000US-0250291
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251188
PR	05-DEC-2000	2000US-0256119

PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Barash SC, Ruben SM;
DR	
XX	WPI: 2001-483426/52.
DR	N-PSDB; AAK62667.
XX	
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating cancers and
PT	metastasis -
XX	
PS	Claim 11: SEQ ID NO 17479; 3071bp + Sequence Listing; English.
XX	
CC	AAK4951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC	amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/hematopoietic-related diseases, especially
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC	to AAK7694 represent human immune/hematopoietic-derived cells. AAK64703
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC	represent sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 97 AA:
Query Match	50.6%; Score 39; DB 22; Length 97;
Best Local Similarity	70.0%; Pred. No. 23;
Matches 7; Conservative	2; Mismatches 1; Indels 0; Gaps 0;
Oy	2 LPMSPRGVIA 11 :::
Db	54 IPLSPAGVIA 63



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:03:04 ; Search time 34 Seconds

(without alignments)  
13.846 Million cell updates/sec

Title: US-09-824-053-3

Perfect score: 77

Sequence: 1 DLPMSRGRVIASTLXF 16

Gaping table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	97.4	16	4	US-08-669-304-3
2	75	97.4	546	4	US-08-669-304-31
3	74	96.1	16	4	US-08-669-304-10
4	39	50.6	468	4	US-09-292-097-2
5	39	50.6	770	1	US-08-416-581B-1
6	39	50.6	770	1	US-08-416-581B-5
7	39	50.6	770	4	US-09-087-465-6
8	39	50.6	771	1	US-08-276-099A-14
9	39	50.6	771	1	US-08-781-890-14
10	38	49.4	393	4	US-09-387-418A-10
11	38	49.4	770	1	US-08-369-796-12
12	38	49.4	770	1	US-08-416-581B-9
13	38	49.4	770	2	US-08-852-091-12
14	38	49.4	770	2	US-08-820-754-12
15	38	49.4	770	3	US-08-956-652-12
16	38	49.4	770	3	US-08-956-869-12
17	38	49.4	770	3	US-09-012-710-8
18	38	49.4	770	3	US-08-948-547-12
19	38	49.4	770	4	US-09-364-970-3
20	38	49.4	770	4	US-09-364-970-5
21	38	49.4	770	4	US-09-556-273-8
22	38	49.4	770	4	US-08-956-653A-12
23	38	49.4	770	4	US-09-526-542-2
24	38	49.4	770	5	PCR-US95-17025-12
25	38	49.4	4544	2	US-08-469-486-52
26	38	49.4	4544	2	US-08-469-658-52
27	37	48.1	307	4	US-08-858-207A-481

28	37	48.1	607	2	US-08-472-534-5	Sequence 5, Appl1
29	37	48.1	616	4	US-09-134-001C-3646	Sequence 3646, Ap
30	37	48.1	861	1	US-08-484-105-18	Sequence 18, Appl
31	37	48.1	861	1	US-08-484-106-18	Sequence 18, Appl
32	37	48.1	2584	3	US-08-936-135-4	Sequence 4, Appl1
33	37	48.1	2588	3	US-08-936-135-2	Sequence 4, Appl1
34	36	46.8	366	4	US-09-134-001C-4799	Sequence 4799, Ap
35	36	46.8	564	4	US-09-211-704A-8	Sequence 8, Appl1
36	36	46.8	669	4	US-08-704-711A-3	Sequence 3, Appl1
37	36	46.8	669	4	US-09-521-220-3	Sequence 3, Appl1
38	36	46.8	669	4	US-09-391-104-29	Sequence 29, Appl
39	35	45.5	298	2	US-09-006-535-3	Sequence 3, Appl1
40	35	45.5	317	1	US-08-118-270-27	Sequence 27, Appl
41	35	45.5	317	5	PCR-US93-08528-27	Sequence 27, Appl
42	34	44.2	31	3	US-08-926-842B-58	Sequence 58, Appl
43	34	44.2	241	3	US-08-834-776A-2	Sequence 2, Appl1
44	34	44.2	244	1	US-08-289-699A-3	Sequence 3, Appl1
45	34	44.2	244	2	US-08-878-285-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-669-304-3  
Sequence 3, Application US/08669304  
Patent No. 6251626  
GENERAL INFORMATION:  
APPLICANT: Peter Strougard  
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A  
TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hunton & Williams  
STREET: 1900 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006-1109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,304  
FILING DATE: 12 July 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,910  
FILING DATE: 7 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stanislaus Akman  
REGISTRATION NUMBER: 28,562  
REFERENCE/DOCKET NUMBER: 320,000003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 778-2201  
TELEFAX: (202) 778-2201  
TELEX: No. 6251626e  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-669-304-3

Query Match 97.4%, Score 75; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 3; 1e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0;

OY 1 DLPMSPRGVIASNLXF 16  
|||||  
Db 1 DLPMSPRGVIASNLXF 16

## RESULT 2

US-08-669-304-31  
Sequence 31, Application US/08669304  
Patent No. 6251626

GENERAL INFORMATION:

APPLICANT: Peter Stougaard

APPLICANT: Ole Cai Hansen

TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A

TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hunton & Williams

STREET: 1900 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006-1109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/669,304

FILING DATE: 12 July 1996

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: 08/476,910

APPLICATION NUMBER: 7 June 1995

FILING DATE: 7 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: Stanislaus Aksman

REGISTRATION NUMBER: 28,562

REFERENCE/DOCKET NUMBER: 320.000003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 955-1926

TELEFAX: (202) 778-2201

TELEX: No. 6251626e

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 546 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-669-304-31

Query Match 97.4%; Score 75; DB 4; Length 546;  
Best local Similarity 93.8%; Pred. No. 1.8e-05;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 DLPMSPRGVIASNLXF 16  
|||||  
Db 219 DLPMSPRGVIASNLHF 234

## RESULT 3

US-08-669-304-10  
Sequence 10, Application US/08669304  
Patent No. 6251626

GENERAL INFORMATION:

APPLICANT: Peter Stougaard

APPLICANT: Ole Cai Hansen

TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A

TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hunton & Williams

STREET: 1900 K Street, N.W.

CITY: Washington

STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006-1109  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/669,304

FILING DATE: 12 July 1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,910

FILING DATE: 7 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: Stanislaus Aksman

REGISTRATION NUMBER: 28,562

REFERENCE/DOCKET NUMBER: 320.000003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 955-1926

TELEFAX: (202) 778-2201

TELEX: No. 6251626e

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-669-304-10

Query Match 96.1%; Score 74; DB 4; Length 16;  
Best local Similarity 93.8%; Pred. No. 4.6e-07;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSPRGVIASNLXF 16  
|||||  
Db 1 DLPMSPRGVIASNLHF 16

## RESULT 4

US-09-292-097-2  
Sequence 2, Application US/09292097B

Patent No. 6322977

GENERAL INFORMATION:

APPLICANT: Ial, Preeti

APPLICANT: Kaser, Matthew, R.

APPLICANT: Baughn, Mariah, R.

TITLE OF INVENTION: TAPASIN-LIKE PROTEIN

FILE REFERENCE: PC-0002 US

CURRENT APPLICATION NUMBER: US/09/292,097B

CURRENT FILING DATE: 1999-04-14

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PERL Program

SEQ ID NO 2

LENGTH: 468

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE: -

OTHER INFORMATION: 103348CD1

US-09-292-097-2

Query Match 50.6%; Score 39; DB 4; Length 468;  
Best local Similarity 37.5%; Pred. No. 39;

Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 DLPMSPRGVIASNLXF 16  
:|:|:|:|:|:|  
Db 176 NLPLSPQGVTRTAVWF 191

## RESULT 5



US-08-416-581B-1  
; Sequence 1, Application US/08416581B  
; Patent No. 5719042  
; GENERAL INFORMATION:  
; APPLICANT: Kishimoto, Tadamitsu  
; APPLICANT: Akira, Shizuo  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR APRE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/416,581B  
; FILING DATE: 04-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-65825/1994  
; FILING DATE: 04-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Nakamura, Dean H.  
; REGISTRATION NUMBER: 33,981  
; REFERENCE/DOCKET NUMBER: O-37891  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)293-7060  
; TELEFAX: (202)293-7860  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 770 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-416-581B-1  
; Query Match 50.6%; Score 39; DB 1; Length 770;  
; Best Local Similarity 77.8%; Pred. No. 70;  
; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
; 1 DLPMSPRGV 9  
; ||||| :  
; 723 DLPMSPRAL 731  
; RESULT 6  
; US-08-416-581B-5  
; Sequence 5, Application US/08416581B  
; Patent No. 5719042  
; GENERAL INFORMATION:  
; APPLICANT: Kishimoto, Tadamitsu  
; APPLICANT: Akira, Shizuo  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR APRE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/416,581B  
; FILING DATE: 04-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-65825/1994  
; FILING DATE: 04-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Nakamura, Dean H.  
; REGISTRATION NUMBER: 33,981  
; REFERENCE/DOCKET NUMBER: O-37891  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)293-7060  
; TELEFAX: (202)293-7860  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 770 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-416-581B-5  
; Query Match 50.6%; Score 39; DB 1; Length 770;  
; Best Local Similarity 77.8%; Pred. No. 70;  
; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
; 1 DLPMSPRGV 9  
; ||||| :  
; 723 DLPMSPRAL 731  
; RESULT 7  
; US-09-824-053-6  
; Sequence 6, Application US/09087465A  
; Patent No. 6160092  
; GENERAL INFORMATION:  
; APPLICANT: Vinkemeier, Uwe  
; APPLICANT: Chen, Xiaomai  
; APPLICANT: Darneil Jr., James E  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF  
; TITLE OF INVENTION: USE  
; FILE REFERENCE: 600-1-229  
; CURRENT APPLICATION NUMBER: US/09/087,465A  
; CURRENT FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 770  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-824-053-6  
; Query Match 50.6%; Score 39; DB 4; Length 770;  
; Best Local Similarity 77.8%; Pred. No. 70;  
; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
; 1 DLPMSPRGV 9  
; ||||| :  
; 723 DLPMSPRAL 731  
; RESULT 8  
; US-08-276-099A-14  
; Sequence 14, Application US/08276099A  
; Patent No. 5591825  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L  
; APPLICANT: Hou, Jinhao  
; TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
; TITLE OF INVENTION: BINDING ASSAYS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
City: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/276,099A  
FILING DATE: 15-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59451-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 771 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-276-099A-14

Query Match 50.6%; Score 39; DB 1; Length 771;  
Best Local Similarity 77.8%; Pred. No. 70;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSRGV 9  
Db 723 DLPMSRPL 731

RESULT 9  
US-08-781-890-14  
Sequence 14, Application US/08781890  
Patent No. 5710266  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L  
APPLICANT: Hou, Jinzhao  
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
NUMBER OF SEQUENCES: 17  
TITLE OF INVENTION: BINDING ASSAYS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
City: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,890  
FILING DATE: 05-JAN-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/276,099  
FILING DATE: 15-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard Aron  
REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59451-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 771 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-781-890-14

Query Match 50.6%; Score 39; DB 1; Length 771;  
Best Local Similarity 77.8%; Pred. No. 70;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 DLPMSRGV 9  
Db 723 DLPMSRPL 731

RESULT 10  
US-09-387-418A-10  
Sequence 10, Application US/09387418A  
Patent No. 6391572  
GENERAL INFORMATION:  
APPLICANT: Zhang, Xiaokui  
APPLICANT: Wzieszczyńska, Melissa H  
APPLICANT: Horvath, Curt M  
APPLICANT: Darnell Jr., James E  
TITLE OF INVENTION: METHODS FOR IDENTIFYING MODULATORS OF TRANSCRIPTIONAL ACTIVAT  
FILE REFERENCE: 600-1-253  
CURRENT FILING DATE: 1999-08-31  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-387-418A-10

Query Match 49.4%; Score 38; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLPMSR 7  
Db 346 DLPMSR 352

RESULT 11  
US-08-369-796-12  
Sequence 12, Application US/08369796  
Patent No. 5716622  
GENERAL INFORMATION:  
APPLICANT: James E. Darnell, Jr.  
APPLICANT: Zilong Wen  
APPLICANT: Curt M. Horvath  
APPLICANT: Zhong Zhong  
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL  
NUMBER OF SEQUENCES: 39  
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
City: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/369,796  
FILING DATE: 06-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-369-796-12

Query Match 49.4%; Score 38; DB 1; Length 770;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLPMSR 7  
DB 723 DLPMSR 729

RESULT 12  
US-08-416-581B-9  
Sequence 9, Application US/08416581B  
Patent No. 5719042  
GENERAL INFORMATION:  
APPLICANT: Kishimoto, Tadamltau  
APPLICANT: Akira, Shizuo  
TITLE OF INVENTION: TRANSCRIPTION FACTOR APRE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MIOW, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/416,581B  
FILING DATE: 04-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-65825/1994  
FILING DATE: 04-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Dean H.  
REGISTRATION NUMBER: 33,981  
REFERENCE/DOCKET NUMBER: 0-37891  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-416-581B-9

Query Match 49.4%; Score 38; DB 1; Length 770;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLPMSR 7  
DB 723 DLPMSR 729

RESULT 13  
US-08-852-091-12  
Sequence 12, Application US/08852091  
Patent No. 5883228  
GENERAL INFORMATION:  
APPLICANT: James E. Darnell, Jr.  
APPLICANT: Zilong Wen  
APPLICANT: Curt M. Horvath  
APPLICANT: Zhong Zhong  
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,091  
FILING DATE: 06-MAY-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/369,796  
FILING DATE: 06-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-852-091-12

Query Match 49.4%; Score 38; DB 2; Length 770;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLPMSR 7  
DB 723 DLPMSR 729

RESULT 14  
US-08-820-754-12  
Sequence 12, Application US/08820754  
Patent No. 5976835

GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/820,754  
FILING DATE: 19-MAR-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-652-12

Query Match 49.4%; Score 38; DB 2; Length 770;  
Best Local Similarity 100.0%; Pred. No. 1e+02; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLPMSR 7  
DB 723 DLPMSR 729

RESULT 15  
US-08-956-652-12  
Sequence 12, Application US/08956652  
Patent No. 6013475  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong

APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,652  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-652-12

Query Match 49.4%; Score 38; DB 3; Length 770;  
Best Local Similarity 100.0%; Pred. No. 1e+02; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLPMSR 7  
DB 723 DLPMSR 729

RESULT 16  
US-08-956-869-12  
Sequence 12, Application US/08956869  
Patent No. 6030808  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,869  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/212,185  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-869-12

Query Match  
Best Local Similarity 49.4%; Score 38; DB 3; Length 770;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DLPMSR 7  
|||||||  
723 DLPMSR 729

RESULT 17  
US-09-012-710-8  
Sequence 8, Application US/09012710  
Patent No. 6087478  
GENERAL INFORMATION:  
APPLICANT: Vinkemeter, Uwe  
APPLICANT: Moarefi, Ismail  
APPLICANT: Darnell, Jr., James E.  
APPLICANT: Kurlivan, John  
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,710  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-012-710-8

Query Match  
Best Local Similarity 49.4%; Score 38; DB 3; Length 770;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DLPMSR 7  
|||||||  
723 DLPMSR 729

RESULT 18  
US-08-948-547-12  
Sequence 12, Application US/08948547  
Patent No. 6124118  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,547  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569

FILING DATE: 19-MAR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/126,588  
 FILING DATE: 24-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jackson Esq., David A.  
 REGISTRATION NUMBER: 26,742  
 REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201 487-5800  
 TELEFAX: 201 343-1684  
 TELEFAX: 133521  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 770 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-948-547-12

Query Match 49.4%; Score 38; DB 3; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLPMSPR 7  
 DB 723 DLPMSPR 729

RESULT 19  
 US-09-364-970-3  
 Sequence 3, Application US/09364970  
 Patent No. 6235873  
 GENERAL INFORMATION:  
 APPLICANT: Bromberg, Jacqueline  
 TITLE OF INVENTION: CONSTITUTIVELY ACTIVE TRANSCRIPTION FACTORS AND THEIR  
 TITLE OF INVENTION: USES FOR IDENTIFYING MODULATORS OF ACTIVITY INCLUDING  
 TITLE OF INVENTION: DYSPROLIFERATIVE CELLULAR CHANGES  
 FILE REFERENCE: 600-1-252  
 CURRENT APPLICATION NUMBER: US/09/364,970  
 CURRENT FILING DATE: 1999-07-31  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 3  
 LENGTH: 770  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-364-970-3

Query Match 49.4%; Score 38; DB 4; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLPMSPR 7  
 DB 723 DLPMSPR 729

RESULT 20  
 US-09-364-970-5  
 Sequence 5, Application US/09364970  
 Patent No. 6235873  
 GENERAL INFORMATION:  
 APPLICANT: Bromberg, Jacqueline  
 TITLE OF INVENTION: CONSTITUTIVELY ACTIVE TRANSCRIPTION FACTORS AND THEIR  
 TITLE OF INVENTION: USES FOR IDENTIFYING MODULATORS OF ACTIVITY INCLUDING  
 TITLE OF INVENTION: DYSPROLIFERATIVE CELLULAR CHANGES  
 FILE REFERENCE: 600-1-252  
 CURRENT APPLICATION NUMBER: US/09/364,970  
 CURRENT FILING DATE: 1999-07-31  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 5

LENGTH: 770  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-364-970-5  
 Query Match 49.4%; Score 38; DB 4; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLPMSPR 7  
 DB 723 DLPMSPR 729  
 Search completed: January 2, 2003, 12:05:31  
 Job time : 36 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:03:04 ; Search time 16 Seconds  
(without alignments)  
96.134 Million cell updates/sec

Title: US-09-824-053-3

Perfect score: 77  
Sequence: 1 DLPMSPRGVASNLXF 16

Listing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Decoded: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	55.8	184	2	E98229
2	43	55.8	184	2	A13056
3	42	54.5	416	2	B65171
4	41	53.2	416	2	E91207
5	41	53.2	416	2	G86053
6	40	51.9	295	2	A60131
7	40	51.9	407	2	G84309
8	39	50.6	113	2	I51025
9	39	50.6	206	2	A64339
10	39	50.6	770	2	A54444
11	38	49.4	272	1	C64009
12	38	49.4	328	2	E95421
13	38	49.4	334	2	E86597
14	38	49.4	334	2	G72024
15	38	49.4	363	2	G72854
16	38	49.4	479	2	S72582
17	38	49.4	670	2	A75542
18	38	49.4	684	2	G90473
19	38	49.4	770	2	I49508
20	38	49.4	1305	2	T31096
21	38	49.4	4543	1	A53102
22	38	49.4	4544	1	S02392
23	38	49.4	4545	1	S25111
24	37	48.1	196	2	T50018
25	37	48.1	245	2	AB1807
26	37	48.1	362	2	T41785
27	37	48.1	368	2	G75425
28	37	48.1	485	1	F71275
29	37	48.1	500	2	C84791

30	37	48.1	605	2	I39837	dnak-type molecule
31	37	48.1	607	2	B86744	dnak protein (impo
32	37	48.1	607	2	S39342	dnak-type molecule
33	37	48.1	607	2	B95060	dnak protein (impo
34	37	48.1	607	2	G97928	hypothetical prote
35	37	48.1	611	2	S09500	dnak-type molecule
36	37	48.1	614	2	B38188	class 1 heat-shock
37	37	48.1	631	2	S71508	dnak-type molecule
38	37	48.1	637	2	B83052	dnak protein PA476
39	37	48.1	749	2	A99438	hypothetical prote
40	37	48.1	860	2	A43920	nucleolar 100k pol
41	37	48.1	861	2	G02329	replication contro
42	37	48.1	885	2	JH0796	nucleolar 100k pol
43	37	48.1	1199	1	S76349	transcription-repa
44	37	48.1	1477	2	T18534	protein-tyrosine k
45	37	48.1	2218	2	B84683	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

E98229

C:Species: Agrobacterium tumefaciens (strain C

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002

C:Accession: E98229

R:Goodner, B.; Hinkle, G.; Gatlung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldm

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: E98229

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-184 <KUR>

A:Cross-references: GB:AE007870; PIDN:AKK89359.1; PID:g15159207; GSPDB:GN00170

C:Genetics:

A:Gene: AGR\_L-1572

A:Map position: linear chromosome

C:Superfamily: Corynebacterium sp. sarcosine oxidase gamma chain

Query Match

Best Local Similarity 55.8%; Score 43; DB 2; Length 184;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPMSPRGVIANLXF 16

DB 54 LPTSPKGVTSGLRF 68

RESULT 2

A13056

sarcosine oxidase gamma subunit [imported] - Agrobacterium tumefaciens (strain C58, D

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002

C:Accession: A13056

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: A13056

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-184 <KUR>

A:Cross-references: GB:AE006889; PIDN:AAL4871.1; PID:g17742519; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: soxg

## RESULT 5

R:Ng, W.V.; Kennedy, S.P.; Maitiras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky  
Jung, K.H.; Alam, M.; Freilts, T.  
Leithner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja



Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: G84309  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-407 <STO>  
A:Cross-references: GB:A8004437; NID:g10581047; PIDN:AG19843.1; GSPDB:GN00138  
C:Gene: VNG1564H

Query Match 51.9%; Score 40; DB 2; Length 407;  
Best Local Similarity 46.2%; Pred. No. 27;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
1 DLPMSPRGVASNN 13  
:|||||:  
223 DVALTPRGVAAAS 235

RESULT 8  
IS1025  
transcription factor - eastern newt (fragment)  
C:Species: *Notophthalmus viridescens*, *Triturus viridescens* (eastern newt)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: IS1025  
R:Simon, H.G.; Nelson, C.; Goff, D.; Laufer, E.; Morgan, B.A.; Tablin, C.  
Dev. Dyn. 202, 1-12, 1995  
A:Title: Differential expression of myogenic regulatory genes and Max-1 during dedifferentiation  
A:Reference number: IS1024; MUID:95218226; PMID:7703517  
A:Accession: IS1025  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-113 <STM>  
A:Cross-references: EMBL:X82837; NID:g577786; PIDN:CAA58044.1; PID:g577787  
C:Gene: Max-1  
A:Gene: Max-1  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:1-45/Domain: homeobox homology (fragment) <HOX>

Query Match 50.6%; Score 39; DB 2; Length 113;  
Best Local Similarity 40.0%; Pred. No. 9.6;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
2 LPMSPRGVASNNLXF 16  
:|||||:  
93 MPMSPKLVAAHVGY 107

RESULT 9  
A64339  
hypothetical protein M0312 - *Methanococcus jannaschii*  
C:Species: *Methanococcus jannaschii*  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: A64339  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64339  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-206 <BU>  
A:Cross-references: GB:U67486; GB:L77117; NID:g1591031; PIDN:AA96309.1; PID:g1592281; T  
C:Genetics:  
A:Map position: REV291845-291225

Query Match 50.6%; Score 39; DB 2; Length 206;

Best Local Similarity 35.7%; Pred. No. 19;  
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
QY 1 DLPMSPRGVASNNL 14  
:|||||:  
DB 153 EFPISPEGLIAASI 166

RESULT 10  
A54444  
DNA-binding protein APRF - human  
C:Species: *Homo sapiens* (man)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 28-Jul-2000  
C:Accession: A54444  
R:Akira, S.; Nishio, Y.; Inoue, M.; Wang, X.J.; Wei, S.; Matsusaka, T.; Yoshida, K.; Cell 77, 63-71, 1994  
A:Title: Molecular cloning of APRF, a novel IFN- $\gamma$ -stimulated gene factor 3 p91-related  
A:Reference number: A54444; MUID:94208062; PMID:7512451  
A:Accession: A54444  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-770 <RES>  
A:Cross-references: GB:L92277; NID:g475788; PID:g475789  
C:Gene: APRF  
A:Gene: GDB:STAT3; APRF  
A:Cross-references: GDB:358950  
A:Map position: 17g21-17g21  
C:Superfamily: human signal transducer and transcription activator STAT3A  
C:Keywords: DNA binding; transcription factor

Query Match 50.6%; Score 39; DB 2; Length 770;  
Best Local Similarity 77.8%; Pred. No. 84;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DLPMSPRGV 9  
:|||||:  
DB 723 DLPMSPRAL 731

RESULT 11  
C64009  
ribonuclease homolog H10526 - *Haemophilus influenzae* (strain Rd KW20)  
C:Species: *Haemophilus influenzae*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C64009  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.H.; Smit, A.M.; Venter, A.; Smith, T.F.; Wilson, R.; Luv, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fritchman, J.L.; Geoghegan, N.S.M.; Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: C64009  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-272 <RTG>  
A:Cross-references: GB:U32735; GB:L42023; NID:g1573509; PIDN:ACC22192.1; PID:g1573510  
C:Superfamily: Enterobacter ribonuclease

Query Match 49.4%; Score 38; DB 1; Length 272;  
Best Local Similarity 50.0%; Pred. No. 39;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 DLPMSPRGVASNNL 14  
:|||||:  
DB 170 DLPALPRGLAAYL 183

RESULT 12  
E95421  
conserved hypothetical protein SKA2361 [Imported] - *Sinorhizobium meliloti* (strain 10  
C:Species: *Sinorhizobium meliloti*  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001



C:Accession: S72582  
 R:Smith, D.R.; Robison, K.  
 submitted to the EMBL Data Library, November 1993  
 A:Reference number: S72580  
 A:Accession: S72582  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-479 <SMI>  
 A:Cross-references: EMBL:U00016; NID:g466961; PIDN:AAAI7150.1; PID:g466964  
 C:Genetics:  
 A:Start codon: GTG

Query Match 49.4%; Score 38; DB 2; Length 479;  
 Best Local Similarity 70.0%; Pred. No. 75;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 2 LPMSPRGVIA 11  
 ||| |||||  
 225 LPLGPRGVNA 234

RESULT 17  
 A75542  
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: A75542  
 R:White, O.; Eisen, J.A.; Heldelberg, J.F.; Hickay, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: A75542  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-670 <WHI>  
 A:Cross-references: GB:AE00186; GB:AE000513; NID:g6457921; PIDN:AAF09837.1; PID:g645792  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0250  
 A:Map position: 1

Query Match 49.4%; Score 38; DB 2; Length 670;  
 Best Local Similarity 58.3%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 5 SPRGVASNLKF 16  
 :|||:|:|:  
 555 TPAGVVASLIF 566

RESULT 18  
 G90473  
 hypothetical protein cutA-6 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-Aug-2001  
 C:Accession: G90473  
 R:Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
 arett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: G90473  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-684 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13816315; PIDN:AAK43046.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: cutA-6  
 C:Superfamily: carbon-monoxide dehydrogenase large chain

Query Match 49.4%; Score 38; DB 2; Length 684;  
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 3 PMSPRGVIA 11  
 ||| |||||  
 176 PMEPKGIIA 184

RESULT 19  
 I49508  
 ISGF3 p91-related transcription factor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 01-Dec-2000  
 C:Accession: I49508; MUID:149009  
 R:Kitt, S.; Nishio, Y.; Inoue, M.; Wang, X.J.; Wei, S.; Matsusaka, T.; Yoshida, K.;  
 Cell 77, 63-71, 1994  
 A:Title: Molecular cloning of APRF, a novel IFN-stimulated gene factor 3 p91-related  
 A:Reference number: A54444; MUID:94208062; PMID:7512451  
 A:Accession: I49508  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-770 <RES>  
 A:Cross-references: GB:L29278; NID:g476715; PIDN:AAA37254.1; PID:g476716  
 R:Raz, R.; Durbin, J.E.; Levy, D.E.  
 J. Biol. Chem. 269, 24391-24395, 1994  
 A:Title: Acute phase response factor and additional members of the interferon-stimula  
 A:Reference number: I49009; MUID:95014185; PMID:7523373  
 A:Accession: I49009  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-393; MUID:395-700; 702-770 <RE2>  
 A:Cross-references: EMBL:U08378; NID:g473889; PIDN:AAA56668.1; PID:g473890  
 C:Genetics:  
 A:Gene: APRF  
 C:Superfamily: human signal transducer and transcription activator STAT5A

Query Match 49.4%; Score 38; DB 2; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 DLPMSPR 7  
 |||||  
 723 DLPMSPR 729

RESULT 20  
 T31096  
 cyclin G-associated kinase GAK - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T31096  
 R:Kanaoka, Y.; Kimura, S.H.; Okazaki, I.; Ikeda, M.; Nojima, H.  
 FEBS Lett. 402, 73-80, 1997  
 A:Title: GAK: a cyclin G-associated kinase contains a tensin/auxilin-like domain.  
 A:Reference number: Z20979; MUID:97165969; PMID:9013862  
 A:Accession: T31096  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1305 <KAN>  
 A:Cross-references: EMBL:D38560; NID:g1902912; PIDN:BAI8911.1; PID:g1902913  
 A:Note: GAK and cyclin G associate together in vivo  
 C:Genetics:  
 A:Gene: GAK

Query Match 49.4%; Score 38; DB 2; Length 1305;  
 Best Local Similarity 61.5%; Pred. No. 2.3e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 1 DLPMSPRGVIA 13  
 ||| |||||  
 1015 DLPMPKVIASS 1027

Thu Jan 2 12:35:48 2003

us-09-824-053-3.rpt

Page 6

Search completed: January 2, 2003, 12:04:49  
Job time : 18 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:03:00 ; Search time 32 Seconds

(without alignments)  
20.738 Million cell updates/sec

Title: US-09-824-053-3  
Sequence: 77  
1 DLPMSRGVIANSLMF 16

Gapop 10.0, Gapext 0.5

112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	54.5	416	1 YIDR_ECOLI	P31455 escherichia
2	40	51.9	295	1 HX71_XENLA	004281 xenopus lae
3	39	50.6	206	1 Y312_METJA	057760 methanococ
4	39	50.6	770	1 STA3_HUMAN	P40763 homo sapien
5	38	49.4	372	1 RN26_HAEIN	P44012 haemophilus
6	38	49.4	334	1 GPDA_CHLPN	Q92751 chlamydia p
7	38	49.4	363	1 VPA3_NPVAC	P34050 autographa
8	38	49.4	590	1 DNK3_STPMV	006942 streptococ
9	38	49.4	770	1 STA3_MOUSE	P42227 mus musculu
10	38	49.4	770	1 STA3_MOUSE	P52631 rattus norv
11	38	49.4	1305	1 GAK_RAT	P97874 rattus norv
12	38	49.4	4543	1 LRP1_CHICK	P98157 gallus gall
13	38	49.4	4544	1 LRP1_HUMAN	Q07954 homo sapien
14	37	48.1	196	1 FLC_ARATH	Q08797 arabidopsis
15	37	48.1	499	1 GAK_MOUSE	Q09KY4 mus musculu
16	37	48.1	500	1 GAPS_PSEAG	Q931Y3 pseudomonas
17	37	48.1	604	1 DNK3_BACME	P05646 bacillus me
18	37	48.1	606	1 DNK3_BACTR	Q9KWS7 bacillus th
19	37	48.1	607	1 DNK3_LACLA	P42368 lactococcus
20	37	48.1	607	1 DNK3_STRPN	P95829 streptococ
21	37	48.1	607	1 DNK3_STRPN	P95831 streptococ
22	37	48.1	609	1 DNK3_STRAG	P95693 streptococ
23	37	48.1	610	1 DNK3_BACSH	P06268 bacillus sp
24	37	48.1	610	1 DNK3_BACSU	P17820 bacillus su
25	37	48.1	613	1 DNK3_BACBD	Q9K472 bacillus ha
26	37	48.1	613	1 DNK3_LACSK	Q08777 lactobacill
27	37	48.1	637	1 DNK3_PSEAE	Q9H443 pseudomonas
28	37	48.1	644	1 DNK3_LEGPN	Q32482 legionella
29	37	48.1	656	1 DNK3_COXBU	Q87312 coxiella bu
30	37	48.1	861	1 ORC1_HUMAN	Q13415 homo sapien
31	37	48.1	885	1 PWC2_HUMAN	Q01780 homo sapien
32	37	48.1	887	1 PWC2_MOUSE	P56660 mus musculu
33	37	48.1	1199	1 MFD_SYNY3	Q55750 synchocyst

34	37	48.1	1377	1 NEOL_RAT	P97603 rattus norv
35	37	48.1	1461	1 NEOL_HUMAN	Q92859 homo sapien
36	37	48.1	1477	1 HTRK_HYDAT	Q25197 hydra atten
37	37	48.1	1493	1 NEOL_MOUSE	P97798 mus musculu
38	37	48.1	1493	1 Y698_CHLMU	Q9PJ88 chlamydia m
39	36.5	47.4	293	1 KC76_MRCTU	Q11043 mycobacteri
40	36	46.8	158	1 NUDC_VIRCH	Q9K427 vibrio chol
41	36	46.8	269	1 PECC_ERWCH	Q07251 erwinia chr
42	36	46.8	349	1 LDH_ALICEU	Q07251 erwinia chr
43	36	46.8	388	1 AMPC_YEREN	P45460 yersinia en
44	36	46.8	409	1 YOPM_YERPE	P17778 yersinia pe
45	36	46.8	421	1 Y118_TREPA	Q83155 treponema p

## ALIGNMENTS

RESULT 1					
YIDR_ECOLI	STANDARD:	PRT:	416 AA.		
ID YIDR_ECOLI	P31455; P76734;				
AC	01-JUL-1993 (Rel. 26, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DE	Hypothetical protein yidr.				
GN	YIDR OR B3689.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12 / MG1655;				
RA	Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;				
RT	"DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication.";				
RL	Genomics 16:551-561(1993).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: L10328; AAA62041.1; -				
DR	EMBL: AEO00446; AAC76712.1; -				
DR	Ecogene; EG11713; YIDR.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 416 AA; 46318 MW; 47E36360CC89A8B7 CRC64;				
-----					
Query Match					
Best Local Similarity		54.5%	Score 42;	DB 1;	Length 416;
Matches		8;	Conservative	1;	Mismatches 6;
					Indels 0;
					Gaps 0;
-----					
QY	2 LPMSPRGVIANSLMF 16				
Db	272 LPMSPRGVIANSLMF 286				
-----					
RESULT 2					
HX71_XENLA	STANDARD:	PRT:	295 AA.		
ID HX71_XENLA	C004281;				
AC	01-JUN-1994 (Rel. 29, Created)				
DT	15-JUN-1994 (Rel. 29, Last sequence update)				
DE	Homeobox protein XHOX-7.1 (Fragment).				
GN	XHOX-7.1.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 CC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91347929; PubMed=1679007;  
 RA Su M.-W., Suzuki H.R., Solursh M., Ramirez F.;  
 RT "Progressively restricted expression of a new homeobox-containing  
 RL gene during xenopus laevis embryogenesis.";  
 CC Development 111:1179-1187(1991).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- DEVELOPMENTAL STAGE: APPEARS AT THE BEGINNING OF GASTRULATION.  
 CC PLATEAU BETWEEN THE NEURULA AND MIDDLE-TAILBUD STAGES, AND  
 CC DECREASE STEADILY THEREAFTER.  
 CC -1- SIMILARITY: BELONGS TO THE MSH HOMEBOX FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: X58773; CAA41574.1; -  
 DR PIR: A60131; A60131.  
 DR HSSP: P22808; 1NK3.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; Homeobox.1.  
 DR ProDom: PD00010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS00071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT NON\_TER 1  
 FT DNA\_BIND 170 HOMEBOX.  
 SO SEQUENCE 295 AA; 32338 MW; 054331E2BC106C10 CRC64;  
 Query Match 51.9%; Score 40; DB 1; Length 295;  
 Best Local Similarity 46.7%; Pred. No. 7.8;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 OY 2 LPMSRGVYASNLX 16  
 DB 275 LPMSRGVYASNLX 289  
 RESULT 3  
 Y312.METJA STANDARD; PRT; 206 AA.  
 AC 057760;  
 RX 01-NOV-1997 (Rel. 35, Created)  
 RX 01-NOV-1997 (Rel. 35, Last sequence update)  
 RX 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN Hypothetical protein M0312.  
 GN M0312.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fudman J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U67486; AAB98309.1; -  
 DR DR TIGR: M0312;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 206 AA; 22367 MW; 1168A5765FE033A CRC64;  
 Query Match 50.6%; Score 39; DB 1; Length 206;  
 Best Local Similarity 35.7%; Pred. No. 7.9;  
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 DLPMSPRGVYASNL 14  
 DB 153 EPLSPGGLAAST 166  
 RESULT 4  
 Y3A3.HUMAN STANDARD; PRT; 770 AA.  
 AC P40763;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Signal transducer and activator of transcription 3 (Acute-phase  
 DE response factor).  
 GN SFR3 OR APRF.  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=94208062; PubMed=7512451;  
 RA Akira S., Nishio Y., Inoue M., Wang X.-J., Wei S., Matsusaka T.,  
 RA Yoshida K., Sudo T., Naruto M., Kishimoto T.;  
 RT "Molecular cloning of APRF, a novel IFN-stimulated gene factor 3 p91-  
 RT related transcription factor involved in the gp130-mediated signaling  
 RT pathway.";  
 RL Cell 77:63-71(1994).  
 RL [2]  
 RP PHOSPHORYLATION ON SERINE.  
 RX MEDLINE=95215843; PubMed=7701321;  
 RA Zhang X., Blenis J., Li H.-C., Schindler C., Chen-Kiang S.;  
 RT "Requirement of serine phosphorylation for formation of STAT-promoter  
 RT complexes.";  
 RL Science 267:1990-1994(1995).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE INTERLEUKIN-6  
 CC (IL-6)-RESPONSIVE ELEMENTS IDENTIFIED IN THE PROMOTERS OF VARIOUS  
 CC ACUTE-PHASE PROTEIN GENES.  
 CC -1- PATHWAY: INVOLVED IN THE Gp130-MEDIATED SIGNALING PATHWAY.  
 CC -1- SUBUNIT: FORMS A HOMODIMER OR A HETERODIMER WITH A RELATED FAMILY  
 CC MEMBER (AT LEAST STAT1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear; translocated into the nucleus in  
 CC response to phosphorylation.  
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL  
 CC MUSCLE, KIDNEY, AND PANCREAS.  
 CC -1- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-6, IL-11, CNTF,  
 CC LIF, CSF-1, EGF, PDGF, IFN-ALPHA AND OSM. SERINE PHOSPHORYLATION  
 CC IS IMPORTANT FOR THE FORMATION OF STABLE DNA-BINDING STARS  
 CC HOMODIMERS AND MAXIMAL TRANSCRIPTIONAL ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE STAT FAMILY OF TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL: L29277; AAS58374.1; -  
 DR HSSP: P42227; IBL1.  
 DR TRANSFAC: T01493; -  
 DR Genew: HGNC:11364; STAT3.  
 DR MIM: 102582; -  
 DR InterPro: IPR000980; SH2.  
 DR InterPro: IPR001217; STAT.  
 DR Pfam: PF00017; SH2; 1.  
 DR Pfam: PF01017; STAT; 1.  
 DR Pfam: PF02864; STAT\_bind; 1.  
 DR SMART: SM00252; SH2; 1.  
 DR PROSITE: PS50001; SH2; 1.  
 DR Transcription regulation; DNA-binding; Nuclear protein;  
 DR Phosphorylation; SH2 domain.  
 FT DOMAIN 580 670 SH2.  
 FT MOD\_RES 705 705 PHOSPHORYLATION (BY JAK) (BY SIMILARITY).  
 FT MOD\_RES 727 727 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 770 AA; 88052 MW; A3DCE815B3B5360 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 770;  
 Best Local Similarity 77.8%; Pred. No. 34;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DLPMSRGV 9  
 Db 723 DLPMSRPL 731

RESULT 5  
 RN26\_HAETN STANDARD; PRT; 272 AA.  
 AC P44012;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable ribonuclease HI0526 precursor.  
 GN HI0526.  
 OS Haemophilus influenzae.  
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OS Haemophilus.

NCBI\_TaxID=727;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-Rd / RW20 / ATCC 51907;  
 MEDLINE=95350630; PubMed=7542800;  
 RA Flaischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Klenavay A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodex A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 Science 269:496-512(1995).

CC -1- SIMILARITY: BELONGS TO THE RNASE T2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL: U32735; AAC22192.1; -  
 DR TIGR: HI0526; -  
 DR InterPro: IPR001568; RNase\_T2.  
 DR Pfam: PF00445; Ribonuclease\_T2; 1.  
 DR PROSITE: PS00530; RNase\_T2\_1; 1.  
 DR PROSITE: PS00531; RNase\_T2\_2; 1.  
 KW Hypothetical protein; Hydrolase; Nuclease; Endonuclease; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 272 PROBABLE RIBONUCLEASE HI0526.  
 FT ACT\_SITE 148 148 BY SIMILARITY.  
 FT ACT\_SITE 195 195 BY SIMILARITY.  
 FT ACT\_SITE 199 199 BY SIMILARITY.  
 SO SEQUENCE 272 AA; 31482 MW; 9D50F99C2CEB392 CRC64;

Query Match 49.4%; Score 38; DB 1; Length 272;  
 Best Local Similarity 50.0%; Pred. No. 16;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 DLPMSRGVIAASNL 14  
 Db 170 DLPALPKGLAQYL 183

RESULT 6  
 GPDA\_CHLPN STANDARD; PRT; 334 AA.  
 AC Q92751;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-  
 DE dependent glycerol-3-phosphate dehydrogenase).  
 GN GPDA OR GPDA OR CPN0855 OR CP1014.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OS NCBI\_TaxID=83558;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-CW1029.  
 RA MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 Nat. Genet. 21:385-389(1999).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN-AR39;  
 RA MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Liner K., Weidman J., Kouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin W., Nelson W., Deboy R., Kolonel J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39.";  
 Nucleic Acids Res. 28:1397-1406(2000).  
 [3]  
 SEQUENCE FROM N.A.  
 STRAIN-J138;  
 RA MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kohara S., Nakazawa T.;  
 "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CW1029 from USA.";  
 Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)(+) =  
 CC glycerone phosphate + NAD(P)H.  
 CC -1- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate  
 CC formation.

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE001667; AAD18993.1; -
CC DR EMBL; AE002258; AAF38792.1; -
CC DR EMBL; AP002548; BAA39063.1; -
CC DR TIGR; CP0104; -
CC DR InterPro; IPR001652; NAD_Gly3P_dh.
CC DR Pfam; PF01210; NAD_Gly3P_dh; 1.
CC DR PRINTS; PR00077; GPDHDRGNASE.
CC DR ProDom; PD001649; NAD_Gly3P_dh; 1.
CC DR PROSITE; PS00957; NAD_G3Pdh; 1.
CC Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC SEQUENCE 334 AA; 36161 MW; 4339ED6FFDD7CEE6 CAC64;
CC -----
CC Query Match 49.4%; Score 38; DB 1; Length 334;
CC Best Local Similarity 50.0%; Pred. No. 20;
CC Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
CC -----
CC QY 3 PMSPRGVASNLXF 16
CC |::||::|||
CC Db 48 PLAPNVVISPNLSF 61
CC -----
CC RESULT 7
CC VP43_NPVAC STANDARD; PRT; 363 AA.
CC AC P34050;
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC DT 01-NOV-1995 (Rel. 32, Last annotation update)
CC DE 43 kDa protein.
CC GN P43.
CC OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
CC OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC NCBI_TaxID=46015;
CC OX NCBI [1]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN-HR3;
CC RC MEDLINE=93333214; PubMed=8474157;
CC RX Carstens E.B., Lu A.L., Chan H.L.B.;
CC "Sequence, transcriptional mapping, and overexpression of p47, a
CC baculovirus gene regulating late gene expression.";
CC J. Virol. 67:2513-2520(1993).
CC [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-C6;
CC RX MEDLINE=94303173; PubMed=8030224;
CC RX Ayres M.D., Howard S.C., Kuzio T., Lopez-Ferber M., Possee R.D.;
CC "The complete DNA sequence of Autographa californica nuclear
CC polyhedrosis virus.";
CC RT Virology 202:586-605(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; L07878; AAA16858.1; -
CC DR EMBL; L22858; AAA66669.1; -

```

```

DR   PIR: A45684; A45684.
SQ   SEQUENCE 363 AA; 43490 MW; DCB20B50B078736 CRC64;
      Query Match          49.4%; Score 38; DB 1; Length 363;
      Best Local Similarity 40.0%; Pred. NO. 22;
      Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY   2 LPMSPRGVIASNLXF 16
      :|:|:|:|:|
DB   163 MPYKKKGVIKSNYAF 197

RESULT 8
ID   DNAX_STRMU
AC   006942;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Chaperone protein dnax (Heat shock protein 70) (Heat shock 70 kDa
GN   protein) (HSP70) (Fragment).
OS   Dnax.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxId=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-GS-5;
RX   MEDLINE=97426041; PubMed=9282745;
RA   Jayaraman G.C., Penders J.E., Burne R.A.;
RT   "Transcriptional analysis of the Streptococcus mutans hrca, gripe and
RT   dnaK genes and regulation of expression in response to heat shock and
RT   environmental acidification.";
RL   Mol. Microbiol. 25:329-341(1997).
CC   -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC   -1- INDUCTION: BY HEAT SHOCK.
CC   -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: U78296; AAC45612.1; -.
DR   HSSP: P04475; 2ABP.
DR   InterPro: IPR001023; HSP70.
DR   Pfam: PF00012; HSP70; 1.
DR   Pfdom: PFD000089; HSP70; 1.
DR   PROSITE: PS00297; HSP70_1; 1.
DR   PROSITE: PS00329; HSP70_2; 1.
DR   PROSITE: PS01036; HSP70_3; 1.
KW   Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT   MOD_RES 173 173
FT   NON_TER 590 590
FT   PHOSPHORYLATIOM (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE 590 AA; 63258 MW; 90787ADF08F702B9 CRC64;

OY   1 DLPMSRGVIASNLXF 16
      :|:|:|:|:|
DB   432 DIPAAPRGVQIEVTF 447

Query Match          49.4%; Score 38; DB 1; Length 590;
Best Local Similarity 43.8%; Pred. NO. 38;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

RESULT 9
ID   STRA3_MOUSE
ID   STR3_MOUSE
AC   P42227;
STANDARD;
PRT; 770 AA.

```



DF 01-NOV-1995 (Rel. 32, Created)  
 DF 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Signal transducer and activator of transcription 3 (Acute-phase response factor).  
 DE STAT3 OR APF.  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 154-158; 181-185 AND 632-640.  
 RC STRAIN-BALB/c; TISSUE=Liver;  
 RX MEDLINE=94208062; PubMed=7512451;  
 AKIra S., Nishio Y., Inoue M., Wang X.J., Shi W., Matsusaka T., Yoshida K., Sudo T., Naruto M., Kishimoto T.;  
 "Molecular cloning of APF, a novel IFN-stimulated gene factor 3 p91-related transcription factor involved in the gp130-mediated signaling pathway.";  
 RL Cell 77:63-71(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Thymus;  
 RX MEDLINE=94188718; PubMed=8140422;  
 RA Zhong Z., Wen Z., Darnell J.E. Jr.;  
 "Stat3, a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and Interleukin-6.";  
 RL Science 264:95-98(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95014185; PubMed=7523373;  
 RA Raz R., Durbin J.E., Levy D.E.;  
 "Acute phase response factor and additional members of the interferon-stimulated gene factor 3 family integrate diverse signals from cytokines, interferons, and growth factors.";  
 RL J. Biol. Chem. 269:24391-24395(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A. (STAT3B).  
 RC STRAIN-BALB/c, and C57BL/6; TISSUE=Liver;  
 RX MEDLINE=96016116; PubMed=7568080;  
 RA Schaefer T.S., Sanders L.K., Nathans D.;  
 "Cooperative transcriptional activity of Jun and Stat3 beta, a short form of Stat3.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9097-9101(1995).  
 RN [5]  
 RP PHOSPHORYLATION OF SER-727, AND MUTAGENESIS.  
 RC MEDLINE=95354205; PubMed=7543024;  
 RA Wen Z., Zhong Z., Darnell J.E. Jr.;  
 "Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation.";  
 RL Cell 82:241-250(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF 136-716.  
 RX MEDLINE=98334373; PubMed=9671298;  
 RA Becker S., Groner B., Mueller C.W.;  
 "Three-dimensional structure of the Stat3beta homodimer bound to DNA.";  
 RL Nature 394:145-151(1998).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE INTERLEUKIN-6 (IL-6)-RESPONSIVE ELEMENTS IDENTIFIED IN THE PROMOTERS OF VARIOUS ACUTE-PHASE PROTEIN GENES. STAT3B INTERACTS WITH THE N-TERMINAL PART OF C-JUN TO ACTIVATE SUCH PROMOTERS IN A COOPERATIVE WAY.  
 CC -1- PATHWAY: INVOLVED IN THE GP130-MEDIATED SIGNALING PATHWAY.  
 CC -1- SUBUNIT: FORMS A HOMODIMER OR A HETERODIMER WITH A RELATED FAMILY MEMBER (AT LEAST STAT1) (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear; translocated into the nucleus in response to phosphorylation.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: STAT3A (SHOWN HERE) AND STAT3B; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: STAT3A IS SEEN IN THE LIVER, SPLEEN, AND KIDNEY. STAT3B IS ALSO DETECTED IN THE LIVER, ALTHOUGH IN A MUCH LESS ABUNDANT MANNER.

CC -1- P.TM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-6, IL-11, CNF, LIF, CSE-1, EGF, PDGF, IFN-ALPHA AND OSM. SERINE PHOSPHORYLATION IS IMPORTANT FOR THE FORMATION OF STABLE DNA-BINDING STAT3 HOMODIMERS AND MAXIMAL TRANSCRIPTIONAL ACTIVITY (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE STAT FAMILY OF TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: L29278; AAA37254.1; -  
 CC EMBL: U06922; AAA19452.1; -  
 CC EMBL: U08378; AAA5668.1; -  
 CC EMBL: U30709; AAC52612.1; -  
 CC PDB: 1B61; 13-JAN-99.  
 CC TRANSFAC: T01574; -  
 CC MGD: MGI:103038; Stat3.  
 CC InterPro: IPR000980; SH2.  
 CC InterPro: IPR001217; STAT.  
 CC Pfam: PF00017; SH2; 1.  
 CC Pfam: PF01017; STAT; 1.  
 CC Pfam: PF02864; STAT\_bind; 1.  
 CC Pfam: PF02865; STAT\_prot; 1.  
 CC SMART: SM00252; SH2; 1.  
 CC PROSITE: PS50001; SH2; 1.  
 CC Transcription regulation; DNA-binding; Nuclear protein; SH2 domain; Phosphorylation; Alternative splicing; 3D-structure.  
 KW DOMAIN 580 670  
 FT MOD\_RES 705 705  
 FT MOD\_RES 727 727  
 FT VARSPIC 716 770  
 FT  
 FT VARIANT 701 701  
 FT MUTAGEN 727 727  
 FT  
 FT CONFLICT 16 16  
 FT CONFLICT 25 25  
 FT CONFLICT 394 394  
 FT SEQUENCE 770 AA; 88053 MW; 6C00626711C8012D CRC64;  
 SQ  
 Query Match 49.4%; Score 38; DB 1; Length 770;  
 Best local Similarity 100.0%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLPMSR 7  
 Db 723 DLPMSR 729  
 ID STR3\_RAT  
 AC STR3\_RAT STANDARD: PRT; 770 AA.  
 AC P52631;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Signal transducer and activator of transcription 3.  
 GN STAT3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=96102059; PubMed=8530402;  
 RA Ripberger J.A., Frits S., Richter K., Hocke G.M., Lottspeich F.,

```

FT domain."
RL FEBS Lett. 402:73-80(1997).
CC
CC -1- FUNCTION: Associates with cyclin G and CDK5. Seems to act as an
CC auxilin homolog that is involved in the uncoating of clathrin-
CC coated vesicles by Hsc70 in non-neuronal cells. Expression
CC oscillates slightly during the cell cycle, peaking at G1 (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Localizes to the perinuclear area and to the
CC trans-Golgi network. Also seen on the plasma membrane, probably at
CC focal adhesions (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 TENSIN DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D38560; BAA18911.1; -.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00226; DnaJ; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ANP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00636; DNaJ_1; FALSE_NEG.
DR PROSITE; PS50076; DNaJ_2; 1.
DR Transfaser; Serine/threonine-protein kinase; ATP-binding;
DR Nuclear protein; Endoplasmic reticulum; Cell cycle.
KW ACT_SITE 40 315
FT ACT_SITE 173 173 PROTEIN KINASE.
FT DOMAIN 405 689 BY SIMILARITY.
FT DOMAIN 1241 1305 TENSIN.
SQ SEQUENCE 1305 AA; 143702 MW; 6D36BD38011C44EE CRC64;
Query Match
Best Local Similarity 49.4%; Score 38; DB 1; Length 1305;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 DLPMSRGVIAASN 13
Db 1015 DLPAPMKVIASS 1027
|||||
RESULT 12
LRPI_CHICK
ID LRPI_CHICK STANDARD; PRT; 4543 AA.
AC P98157;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Low-density lipoprotein receptor-related protein 1 precursor (LRP)
DE (Alpha-2-macroglobulin receptor) (A2MR).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Liver, and Ovary;
RX MEDLINE=94103212; PubMed=7506255;
RA Ntmt1 J., Stifani S., Bilous P.T., Schneider W.J.;
```



	RESULT 13
ID	LRP1_HUMAN
AC	Q07954;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Low-density lipoprotein receptor-related protein 1 precursor (LRP1)
DE	(Alpha-2-macroglobulin receptor) (AZMR) (Apolipoprotein E receptor 1)
DE	(ARDR) (CG91).
DE	LRP1 OR A2MR.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	
RX	TISSUE=Liver;
RX	MEDLINE=89210795; PubMed=3266596;
RA	Herz J., Hamann U., Rogne S., Myklebost O., Gausepohl H.,
RA	Stanley K.K.;
RT	"Surface location and high affinity for calcium of a 500-kd liver
RT	membrane protein closely related to the LDL-receptor suggest a
RT	physiological role as lipoprotein receptor.,"
RL	EMBO J. 7:4119-4127(1988).
ET	[2]

```

RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=95203893; PubMed=7534747;
RA Van Leuven F., Stas L., Hilliker C., Lorent K., Umans L., Sernaeels L.,
RA Overbergh L., Torrekens S., Moechars D., De Strooper B.,
RA Van den Berghe H.;
RT "Structure of the gene (LRP1) coding for the human alpha
RT 2-macroglobulin receptor lipoprotein receptor-related protein.";
RL Genomics 24:78-89(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX RX MEDLINE=99000832; PubMed=9782078;
RA Van Leuven F., Stas L., Thyly E., Nellissen B., Miyake Y.;
RT "Strategy to sequence the 65 exons of the human LRP1 gene coding for
RT the lipoprotein receptor related protein: identification of one
RT expressed mutation among 48 polymorphisms.";
RL Genomics 52:138-144(1998).
RN [4]
RP PROCESSING.
RX RX MEDLINE=90269210; PubMed=2112085;
RA Herz J., Kowal R.C., Goldstein J.L., Brown M.S.;
RT "Proteolytic processing of the 600 kd low density lipoprotein
RT receptor-related protein (LRP) occurs in a trans-Golgi
RT compartment.";
RL EMBL J. 9:11769-1776(1990).
RN [5]
RP FUNCTION.
RX RX MEDLINE=91092405; PubMed=1702392;
RA Kristensen T., Moestrup S.K., Gilemann J., Bendtsen L., Sand O.,
RA Sottrup-Jensen L.;
RT "Evidence that the newly cloned low-density-lipoprotein receptor
RT related protein (LRP) is the alpha 2-macroglobulin receptor.";
RL FEBS Lett. 276:151-155(1990).
RN [6]
RP STRUCTURE BY NMR OF 1059-1100
RX RX MEDLINE=99253972; PubMed=10318830;
RA Huang W., Dolmer K., Gettings P.G.W.;
RT "NMR solution structure of complement-like repeat CR8 from the low
RT density lipoprotein receptor-related protein.";
RL Biol. Chem. 274:14130-14136(1999)
RN [7]
RP FUNCTION: INVOLVED IN THE PLASMA CLEARANCE OF CHYLOMICRON REMNANTS
RX RX AND ACTIVATED ALPHA 2-MACROGLOBULIN, AS WELL AS THE LOCAL
RX RX METABOLISM OF COMPLEXES BETWEEN PLASMINOGEN ACTIVATORS AND THEIR
RX RX ENDOGENOUS INHIBITORS.
CC CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN LIVER, BRAIN AND LUNG.
CC CC -1- PHD: CLEAVED INTO A 85 kDa MEMBRANE-SPANNING SUBUNIT (LRP-85) AND
CC CC A 515 kDa LARGE EXTRACELLULAR DOMAIN (LRP-515) THAT REMAINS NON-
CC CC COVALENTLY ASSOCIATED.
CC CC -1- SIMILARITY: CONTAINS 22 EGF-LIKE DOMAINS.
CC CC -1- SIMILARITY: CONTAINS 31 LDL-RECEPTOR CLASS A DOMAINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC CC or send an email to license@isb-slb.ch).
CC CC -----
DR DR EMBL: X13916; CAA32112.1; -
DR DR EMBL: AF058427; AAC64265.1; -
DR DR PDB: 1CR8; 06-JAN-99.
DR DR Genew; HGNC:6692; LRP1.
DR DR MIM: 107770; -
DR DR InterPro: IPR000152; Asx_hydroxyl.
DR DR InterPro: IPR000561; EGF-like.
DR DR InterPro: IPR001881; EGF_Ca.
DR DR InterPro: IPR002172; LDL_recept_A.
DR DR InterPro: IPR000033; ldl_receptor_rep.
DR DR Pfam: PF00008; EGF; 16.
DR DR Pfam: PF00057; ldl_recept_a; 31.
DR DR Pfam: PF00058; ldl_recept_b; 33.
DR DR PRINTS: PR00261; LDLRECEPTOR.

```

DR SMART; SM00179; EGF\_CA; 3.  
 DR SMART; SM00001; EGF\_Like; 19.  
 DR SMART; SM00192; LDla; 31.  
 DR SMART; SM00135; LY; 32.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS00022; EGF\_1; 5.  
 DR PROSITE; PS01186; EGF\_2; 8.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 27.  
 DR PROSITE; PS00068; LDLRA\_2; 31.  
 DR Receptor; Transmembrane; Repeat; Endocytosis; Glycoprotein;  
 KW Signal; Calcium-binding; EGF-like domain; Coated pits; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 4544  
 FT DOMAIN 20 4419  
 FT TRANSMEM 4420 4444  
 FT DOMAIN 4445 4544  
 FT DOMAIN 25 66  
 FT DOMAIN 70 110  
 FT DOMAIN 111 119  
 FT DOMAIN 150 189  
 FT DOMAIN 474 520  
 FT DOMAIN 803 843  
 FT DOMAIN 852 893  
 FT DOMAIN 893 933  
 FT DOMAIN 934 973  
 FT DOMAIN 974 1013  
 FT DOMAIN 1013 1053  
 FT DOMAIN 1060 1099  
 FT DOMAIN 1102 1142  
 FT DOMAIN 1143 1182  
 FT DOMAIN 1183 1222  
 FT DOMAIN 1223 1262  
 FT DOMAIN 1536 1579  
 FT DOMAIN 1846 1887  
 FT DOMAIN 2155 2195  
 FT DOMAIN 2478 2518  
 FT DOMAIN 2522 2563  
 FT DOMAIN 2564 2602  
 FT DOMAIN 2603 2641  
 FT DOMAIN 2642 2690  
 FT DOMAIN 2694 2732  
 FT DOMAIN 2732 2771  
 FT DOMAIN 2772 2814  
 FT DOMAIN 2816 2855  
 FT DOMAIN 2856 2899  
 FT DOMAIN 2902 2940  
 FT DOMAIN 2941 2981  
 FT DOMAIN 2982 3022  
 FT DOMAIN 3290 3331  
 FT DOMAIN 3332 3371  
 FT DOMAIN 3372 3410  
 FT DOMAIN 3411 3450  
 FT DOMAIN 3451 3491  
 FT DOMAIN 3492 3533  
 FT DOMAIN 3534 3572  
 FT DOMAIN 3573 3611  
 FT DOMAIN 3611 3649  
 FT DOMAIN 3652 3692  
 FT DOMAIN 3693 3733  
 FT DOMAIN 3739 3778  
 FT DOMAIN 3781 3823  
 FT DOMAIN 3824 3861  
 FT DOMAIN 4147 4183  
 FT DOMAIN 4196 4232  
 FT DOMAIN 4232 4268  
 FT DOMAIN 4268 4304  
 FT DOMAIN 4304 4340  
 FT DOMAIN 4340 4375  
 FT DOMAIN 4373 4409  
 FT SITE 3940 3943

LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED  
 PROTEIN 1.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 LDL-RECEPTOR CLASS A 1.  
 LDL-RECEPTOR CLASS A 2.  
 EGF-LIKE 1.  
 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 3.  
 EGF-LIKE 4.  
 LDL-RECEPTOR CLASS A 3.  
 LDL-RECEPTOR CLASS A 4.  
 LDL-RECEPTOR CLASS A 5.  
 LDL-RECEPTOR CLASS A 6.  
 LDL-RECEPTOR CLASS A 7.  
 LDL-RECEPTOR CLASS A 8.  
 LDL-RECEPTOR CLASS A 9.  
 LDL-RECEPTOR CLASS A 10.  
 EGF-LIKE 5.  
 EGF-LIKE 6.  
 EGF-LIKE 7.  
 EGF-LIKE 8.  
 EGF-LIKE 9.  
 EGF-LIKE 10.  
 LDL-RECEPTOR CLASS A 11.  
 LDL-RECEPTOR CLASS A 12.  
 LDL-RECEPTOR CLASS A 13.  
 LDL-RECEPTOR CLASS A 14.  
 LDL-RECEPTOR CLASS A 15.  
 LDL-RECEPTOR CLASS A 16.  
 LDL-RECEPTOR CLASS A 17.  
 LDL-RECEPTOR CLASS A 18.  
 LDL-RECEPTOR CLASS A 19.  
 LDL-RECEPTOR CLASS A 20.  
 EGF-LIKE 11.  
 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 13.  
 LDL-RECEPTOR CLASS A 21.  
 LDL-RECEPTOR CLASS A 22.  
 LDL-RECEPTOR CLASS A 23.  
 LDL-RECEPTOR CLASS A 24.  
 LDL-RECEPTOR CLASS A 25.  
 LDL-RECEPTOR CLASS A 26.  
 LDL-RECEPTOR CLASS A 27.  
 LDL-RECEPTOR CLASS A 28.  
 LDL-RECEPTOR CLASS A 29.  
 LDL-RECEPTOR CLASS A 30.  
 LDL-RECEPTOR CLASS A 31.  
 EGF-LIKE 14.  
 EGF-LIKE 15.  
 EGF-LIKE 16.  
 EGF-LIKE 17.  
 EGF-LIKE 18.  
 EGF-LIKE 19.  
 EGF-LIKE 20.  
 EGF-LIKE 21.  
 EGF-LIKE 22.  
 RECOGNITION SITE FOR PROTEOLYTICAL  
 PROCESSING (POTENTIAL).

FT SITE 4502 4507  
 FT DISULFID 27 40  
 FT DISULFID 34 53  
 FT DISULFID 47 64  
 FT DISULFID 72 85  
 FT DISULFID 79 98  
 FT DISULFID 92 108  
 FT DISULFID 115 124  
 FT DISULFID 120 133  
 FT DISULFID 135 148  
 FT DISULFID 154 164  
 FT DISULFID 160 173  
 FT DISULFID 175 188  
 FT DISULFID 478 493  
 FT DISULFID 489 504  
 FT DISULFID 506 519  
 FT DISULFID 807 818  
 FT DISULFID 814 827  
 FT DISULFID 829 842  
 FT DISULFID 854 866  
 FT DISULFID 861 879  
 FT DISULFID 873 890  
 FT DISULFID 895 907  
 FT DISULFID 902 920  
 FT DISULFID 914 931  
 FT DISULFID 936 948  
 FT DISULFID 943 961  
 FT DISULFID 955 971  
 FT DISULFID 976 989  
 FT DISULFID 984 1002

Query Match 49.4%; Score 38; DB 1; Length 4544;  
 Best Local Similarity 54.5%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DLPMSRGVIA 11  
 DB 2459 DLPQPMGIIA 2469

RESULT 14  
 FLC\_ARATH STANDARD; PRT; 196 AA.  
 AC 095707;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE FLOWERING LOCUS C protein (MADS box protein FLOWERING LOCUS F).  
 GN FLC OR FLF OR ATSG10140 OR T31P16.130.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. C24, and cv. Columbia;  
 RX MEDLINE=99172058; PubMed=10072403;  
 RA Sheldon C.C., Burn J.E., Perez P.P., Metzger J., Edwards J.A.,  
 RA Peacock W.J., Dennis E.S.;  
 RT "The FLF MADS box gene: A repressor of flowering in Arabidopsis  
 RT regulated by vernalization and methylation";  
 RU Plant Cell 11:445-456(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99264314; PubMed=10330478;  
 RA Michaels S.D., Amasino R.M.;  
 RT "FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a  
 RT repressor of flowering";  
 RU Plant Cell 11:949-956(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;

RX MEDLINE=21016721; PubMed=11130714;  
 RA Tabata S., Kaneo T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuo A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Naruo K., Okumura S., Shimpo S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Hebermann K., Murray J., Johnson D., Rohlffing T., Nelson J.,  
 RA Stoeneking T., Pepin K., Spiehl J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney M.,  
 RA Du H., Edwards J., Fryman J., Haekensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,  
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,  
 RA Kirschoff K., Toth K., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Marienssen R., McCombie W.R., Wilson R.K., Murphy B., Barcroft I.,  
 RA Volckaert G., Wambut R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Ertan K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
 RA Ramsperger U., Medler H., Balke K., Wedler E., Peters S.,  
 RA van Staveren M., Dirse W., Moolman P., Klein Lankhorst R.,  
 RA Weitzenecker T., Bothe G., Rose M., Hauf J., Bernels S., Hempel S.,  
 RA Feldpausch M., Lambers S., Villarroel R., Gielen J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.,  
 RA "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 thaliana."  
 Nature 408:823-826(2000).  
 [4]  
 RP FUNCTION.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20202708; PubMed=10716723;  
 RA Sheldon C.C., Rouse D.T., Flanagan E.J., Peacock W.J., Dennis E.S.,  
 RT "The molecular basis of vernalization: the central role of FLOWERING  
 LOCUS C (FLC)."  
 Proc. Natl. Acad. Sci. U.S.A. 97:3753-3758(2000).  
 RN [5]  
 RP FUNCTION.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21178818; PubMed=11283346;  
 RA Michaels S.D., Amasino R.M.;  
 RT "Loss of flowering locus c actively eliminates the late-flowering  
 phenotype of *Arabidopsis thaliana* and autonomous pathway mutations but not  
 plant cell 13:935-942(2001)."  
 RL [5]  
 CC -1- FUNCTION: Putative transcription factor that seems to play a  
 CC central role in the regulation of flowering time in the late-  
 CC flowering phenotype by interacting with 'FRIGIDA', the autonomous  
 CC repressing 'SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1',  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN THE VEGETATIVE APEX AND IN  
 CC ROOT TISSUE AND LOWER EXPRESSION IN LEAVES AND STEMS. NOT DETECTED  
 CC IN YOUNG TISSUES OF THE INFLORESCENCE.  
 CC -1- DEVELOPMENTAL STAGE: FOUND IN SHOOTS OF NON-FLOWERING PLANTS GROWN  
 CC UNDER LONG-DAY CONDITIONS AT DAYS 4 TO 15, AND IN SHOOTS OF PLANTS  
 CC GROWN UNDER SHORT-DAY CONDITIONS AT DAYS 4 TO 11 AFTER  
 CC GERMINATION.  
 CC -1- INDUCTION: EPIGENETICALLY DOWNREGULATED BY VERNALIZATION  
 CC FACTORS.  
 CC -1- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION  
 CC FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 K-BOX DIMERIZATION DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: AF116528; AAD21249.1; -  
 DR EMBL: AF116527; AAD21248.1; -

DR EMBL: AL356332; CAB92055.1; -  
 DR HSSP: P11746; 1NMW.  
 DR TRANSFAC: T03284.  
 DR InterPro: IPR002487; TF\_Kbox.  
 DR InterPro: IPR002100; TF\_MADSbox.  
 DR Pfam: PF00319; SNF-TF; 1.  
 DR Pfam: PF01486; K-box; 1.  
 DR PRINTS: PR00404; MADSBOX.  
 DR SMART: SM00432; MADS. 1.  
 DR PROSITE: PS00350; MADS\_BOX\_1; FALSE\_NEG.  
 DR PROSITE: PS00066; MADS\_BOX\_2; 1.  
 KW Developmental regulation; Repressor; DNA-binding; Flowering;  
 FT DOMAIN 3 57  
 FT DOMAIN 68 166  
 FT DOMAIN 166 166  
 SQ SEQUENCE 196 AA; 21865 MW; 656085A2B157FF42 CRC64;  
 Query Match 48.1%; Score 37; DB 1; Length 196;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DPMSPRGVASNL 14  
 DB 175 EMEASPAQISDNL 188  
 RESULT 15  
 GAK\_MOUSE STANDARD; PRT; 499 AA.  
 ID GAK\_MOUSE  
 AC O99KV4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cyclin G-associated kinase (EC 2.7.1.1-) (Fragment).  
 GN GAK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Associates with cyclin G and CDKs. Seems to act as an  
 CC auxilin homolog that is involved in the uncoating of clathrin-  
 CC coated vesicles by Hsc70 in non-neuronal cells. Expression  
 CC oscillates slightly during the cell cycle, peaking at G1 (by  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Localizes to the perinuclear area and to the  
 CC trans-Golgi network. Also seen on the plasma membrane, probably at  
 CC focal adhesions (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: BC003956; AAH03956.1; -  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR InterPro: IPR000719; Ser\_Pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_Pkinase.  
 DR Pfam: PF00226; DnaJ\_1.  
 DR SMART: SM00271; DnaJ\_1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; PARTIAL.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; PARTIAL.  
 DR PROSITE: PS00636; DnaJ\_1; FALSE\_NEG.

DR PROSITE; PS50076; DNAJ\_2; 1.  
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KM Nuclear protein; Endoplasmic reticulum; Cell cycle.  
 FT NON TER 1  
 FT DOMAIN 435 499 J-DOMAIN  
 SO SEQUENCE 499 AA; 53175 MW; 12D7D33A92141B96 CRC64;

Query Match 48.1%; Score 37; DB 1; Length 499;  
 Best Local Similarity 61.5%; Pred. No. 48;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 DLPMSRGVIAASN 13  
 DB 209 DLPAPSGVIASS 221

AC Q93JY3; STANDARD; PRT; 500 AA.  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glucosylglycerol-phosphate synthase (EC 2.4.1.213) (Glucosyl-glycerol-phosphate synthase) (GG-PPS)  
 GN GGPS.  
 OS Pseudomonas anguilliseptica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_TaxID=53406;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Hagemann M.; Steindrich R.;  
 RC STRAIN=OAI46;  
 RT "Isolation and characterization of a  
 RT glucosylglycerol-phosphate synthase from heterotrophic bacteria. In  
 RT preparation."  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Involved in salt tolerance by producing GG-phosphate  
 CC from ADP-glucose and glycerol-3-phosphate (G3P), an intermediate  
 CC in the synthesis of the osmolyte glucosylglycerol (GG).  
 CC -1- CATALYTIC ACTIVITY: ADP-glucose + sn-glycerol 3-phosphate -> 2-  
 CC (beta-D-glucosyl)-sn-glycerol 3-phosphate + ADP.  
 CC -1- PATHWAY: Glucosylglycerol biosynthesis; first step.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 20.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; AJ318784; CAC50077.1; -  
 DR InterPro; IPR001830; G1\_20.  
 DR Pfam; PF00982; Glyco-transf\_20; 1.  
 KW Transferrase; Glycosyltransferase.  
 SO SEQUENCE 500 AA; 56849 MW; D0B85214278BD7DA CRC64;

Query Match 48.1%; Score 37; DB 1; Length 500;  
 Best Local Similarity 50.0%; Pred. No. 48;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PMSRGVIAASNLXF 16  
 DB 23 PTPSGIITPLLSF 36

RESULT 17  
 ID DNAJ\_BACME STANDARD; PRT; 604 AA.  
 AC P05646;

DT 01-NOV-1988 (Rel. 09, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNAK.  
 OS Bacillus megaterium.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=87231083; PubMed=3035506;  
 RA Sussman M.D.; Seelow P.;  
 RT "Nucleotide sequence of a Bacillus megaterium gene homologous to the  
 RT dnaK gene of Escherichia coli."  
 RL Nucleic Acids Res. 15:3923-3923(1987).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; Y00154; CAA68348.1; -  
 DR HSP: P04475; IDG4.  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; HSP70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT\_MET 0  
 FT MOD\_RES 172 172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SO SEQUENCE 604 AA; 65119 MW; 29DF856DC5FAC14 CRC64;

Query Match 48.1%; Score 37; DB 1; Length 604;  
 Best Local Similarity 43.8%; Pred. No. 60;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 DLPMSRGVIAASNLXF 16  
 DB 430 DLPAPSGVIAQIEVSF 445

RESULT 18  
 ID DNAJ\_BACTR STANDARD; PRT; 606 AA.  
 AC Q9KWS7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNAK.  
 OS Bacillus thermoglucosidasius.  
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.  
 OX NCBI\_TaxID=1426;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=KP1006;  
 RA Watanabe K.; Iwashiro T.; Suzuki Y.;  
 RT Features of dnaK operon genes of the obligate thermophile Bacillus  
 RT thermoglucosidasius KP1006.  
 RL Antonie Van Leeuwenhoek 77:241-250(2000).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AB017035; BAB03215.1; -  
CC HSP70; P04475; IDG4.  
CC InterPro: IPR001023; HSP70.  
CC Pfam: PF00012; HSP70; 1.  
CC PRINTS: PR00301; HEATSHOCK70.  
CC ProDom: PD000089; HSP70\_1; 1.  
CC PROSITE: PS00297; HSP70\_1; 1.  
CC PROSITE: PS00329; HSP70\_2; 1.  
CC PROSITE: PS01036; HSP70\_3; 1.  
CC Chaperone; ATP-binding; Heat shock; Phosphorylation.  
CC INIT\_MET 0  
CC MOD\_RES 168 168  
CC SEQUENCE 606 AA; 65951 MW; 7955FD3D845F56 CRC64;  
CC  
CC Query Match 48.1%; Score 37; DB 1; Length 606;  
CC Best Local Similarity 43.8%; Pred. No. 60;  
CC Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
CC  
CC QY 1 DLPMSRGVYASNLXP 16  
CC Db 426 DIPAPRGVQIEVTF 441  
CC  
CC RESULT 19  
CC DNK\_LACIA STANDARD: PRT; 607 AA.  
AC P42368; O9CGY8; 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chaperone protein dnk (Heat shock protein 70) (Heat shock 70 kDa  
DE protein) (HSP70).  
GN DNK OR IL0954.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis), and  
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1356; 1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=L. lactis; STRAIN=MG1363;  
RX MEDLINE=94172317; PubMed=8126443;  
RA Eaton T.J., Shearman C.A., Gasson M.J.;  
RT "Cloning and sequence analysis of the dnk gene region of Lactococcus  
RT lactis subsp. lactis";  
RL J. Gen. Microbiol. 139:3253-3263(1993).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC SPECIES=L. lactis; STRAIN=LM0230;  
RX MEDLINE=94237496; PubMed=8181763;  
RA Barril M.J.S., Kim S.G., Batt C.A.;  
RT "Cloning and sequencing of the Lactococcus lactis subsp. lactis dnk  
RT gene using a PCR-based approach";  
RL Gene 142:91-96(1994).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC SPECIES=L. lactis; STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Weissenbach J., Wincker P., Manger S., Jallion O., Malarne K.,  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
RT lactis ssp. lactis IL1403";  
RL Genome Res. 11:731-753(2001).  
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).

CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -----  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X75428; CA554089.1; -  
CC EMBL: X75428; CA553179.1; -  
CC HSP70; P04475; IDG4.  
CC InterPro: IPR001023; HSP70.  
CC Pfam: PF00012; HSP70; 1.  
CC PRINTS: PR00301; HEATSHOCK70.  
CC ProDom: PD000089; HSP70\_1; 1.  
CC PROSITE: PS00297; HSP70\_1; 1.  
CC PROSITE: PS00329; HSP70\_2; 1.  
CC PROSITE: PS01036; HSP70\_3; 1.  
CC Chaperone; ATP-binding; Heat shock; Phosphorylation.  
CC MOD\_RES 173 173  
CC SEQUENCE 607 AA; 64948 MW; 567FA7160333310E CRC64;  
CC  
CC Query Match 48.1%; Score 37; DB 1; Length 607;  
CC Best Local Similarity 37.5%; Pred. No. 60;  
CC Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
CC  
CC QY 1 DLPMSRGVYASNLXP 16  
CC Db 431 DIPAPRGVQIEVTF 446  
CC  
CC RESULT 20  
CC DNK\_STRPN STANDARD: PRT; 607 AA.  
AC P55829; O66035; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chaperone protein dnk (Heat shock protein 70) (Heat shock 70 kDa  
DE protein) (HSP70).  
GN DNK OR SP0517.  
OS Streptococcus pneumoniae.  
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S. pneumoniae; STRAIN=Rx / CP1200;  
RX MEDLINE=98231633; PubMed=9570114;  
RA Kim S.-W., Choi I.-H., Kim S.-N., Kim Y.-H., Pyo S.-N., Rhee D.-K.;  
RT "Molecular cloning, expression, and characterization of dnk in  
RT Streptococcus pneumoniae";  
RL FEMS Microbiol. Lett. 161:217-224(1998).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S. pneumoniae; STRAIN=TTGR4;



RX MEDLINE=21357209; PubMed=11463916;  
RA Tetteil H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heideberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapfel E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RA "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
CC -!- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -!- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC  
CC EMBL; U72720; AAB39221.1; -;  
CC EMBL; U84387; AAC15892.1; -;  
CC EMBL; AE007363; AAK74675.1; -;  
CC HSP; P04475; IDG4.  
CC TIGR; SP0517; -;  
CC InterPro; IPR001023; HSP70.  
CC Pfam; PF00012; HSP70; 1.  
CC PRINTS; PR00301; HEATSHOCK70.  
CC ProDom; PD000089; Hsp70; 1.  
CC PROSITE; PS00297; HSP70\_1; 1.  
CC PROSITE; PS00329; HSP70\_2; 1.  
CC PROSITE; PS01036; HSP70\_3; 1.  
CC Chapterone; ATP-binding; Heat shock; Phosphorylation;  
CC Complete proteome.  
CC  
CC MOD\_RES 173 173 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
CC FT CONFLICT 493 493 A -> P (IN REF. 2).  
CC FT CONFLICT 495 495 A -> S (IN REF. 1).  
CC FT CONFLICT 530 530 A -> P (IN REF. 2).  
CC FT CONFLICT 556 556 T -> A (IN REF. 1 AND 2).  
CC FT CONFLICT 601 602 DG -> E (IN REF. 2).  
CC SQ SEQUENCE 607 AA; 64842 MM; 11D626F1B37D0760 CRC64;  
-----  
Query Match 48.18; Score 37; DB 1; Length 607;  
Best Local Similarity 37.38; Pred. NO. 60;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
OY 1 DLPSPRGVINSNLXF 16  
DB 431 DIPNAPRGIPQIEVTF 446

Search completed: January 2, 2003, 12:03:45  
Job time : 35 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:03:04 ; Search time 100 Seconds

(without alignments)  
32.968 Million cell updates/sec

Title: US-09-824-053-3

Sequence: 1 DLPMSRGVYASNLXF 16

Gaping table: BLOSUM62

Gaped: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: 1: SP archaea:\*

2: SP bacteria:\*

3: SP fungi:\*

4: SP human:\*

5: SP\_invertebrate:\*

6: SP\_mammal:\*

7: SP\_mmc:\*

8: SP\_organelle:\*

9: SP\_phage:\*

10: SP\_plant:\*

11: SP\_fodent:\*

12: SP\_virus:\*

13: SP\_vertebrate:\*

14: SP\_unclassified:\*

15: SP\_virus:\*

16: SP\_bacteriophage:\*

17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

*Applicants should provide*

Result No.	Score	Match	Length	DB ID	Description
1	75	97.4	546	10	P93762
2	44	57.1	1096	11	Q9D504
3	44	57.1	1207	11	Q9D5V1
4	43	55.8	184	16	Q8U8M0
5	43	55.8	631	5	017051
6	41	53.2	416	16	Q8XR29
7	41	53.2	1051	3	P97961
8	40	51.9	229	13	Q90235
9	40	51.9	407	17	Q9HPM4
10	40	51.9	682	17	Q974B5
11	40	51.9	989	5	Q9N9Z1
12	40	51.9	1193	4	Q92580
13	39	50.6	64	10	Q943R9
14	39	50.6	112	10	Q943K8
15	39	50.6	113	13	Q91153
16	39	50.6	150	10	Q8S724

17	39	50.6	280	13	Q91152	Q91152 notophthalm
18	39	50.6	442	16	Q981L6	Q981L6 rhizobium 1
19	39	50.6	468	4	Q9NMB8	Q9NMB8 homo sapien
20	39	50.6	468	4	Q9BX59	Q9BX59 homo sapien
21	39	50.6	541	5	Q9Y1L1	Q9Y1L1 locusta mlg
22	39	50.6	1143	10	Q948Y6	Q948Y6 volvox cart
23	38	49.4	51	4	Q9BXH2	Q9BXH2 homo sapien
24	38	49.4	51	11	Q99ML3	Q99ML3 mus musculu
25	38	49.4	211	2	Q8VW46	Q8VW46 dreivibacter
26	38	49.4	328	16	Q92XK5	Q92XK5 rhizobium m
27	38	49.4	476	5	Q9N8C7	Q9N8C7 trypansom
28	38	49.4	479	16	Q49746	Q49746 mycobacteri
29	38	49.4	534	17	Q97CT0	Q97CT0 thermoplas
30	38	49.4	618	2	Q93R27	Q93R27 tetragenoco
31	38	49.4	622	10	Q941X1	Q941X1 oryza sativ
32	38	49.4	670	16	Q9RX07	Q9RX07 delinococcus
33	38	49.4	684	17	Q9WU00	Q9WU00 sulfolobus
34	38	49.4	769	13	Q9PVM4	Q9PVM4 homo sapien
35	38	49.4	770	13	Q9PVM8	Q9PVM8 xenopus lae
36	38	49.4	890	10	Q8RY23	Q8RY23 arabidopsis
37	38	49.4	3223	2	Q9RPH1	Q9RPH1 escherichia
38	38	49.4	3223	2	Q9RPH8	Q9RPH8 escherichia
39	38	49.4	3223	2	Q8VOR2	Q8VOR2 escherichia
40	38	49.4	3223	2	Q8VNR6	Q8VNR6 escherichia
41	38	49.4	4545	11	Q61291	Q61291 mus musculu
42	38	49.4	4545	11	Q920Y4	Q920Y4 mus musculu
43	38	49.4	4545	11	Q912X7	Q912X7 mus musculu
44	38	49.4	4545	11	Q912X7	Q912X7 mus musculu
45	37	48.1	245	16	Q820S3	Q820S3 anabaena sp

## ALIGNMENTS

RESULT 1	P93762	PRELIMINARY;	PRT;	546 AA.
ID	P93762;			
AC	P93762;			
DT	01-MAY-1997 (TREMBLrel. 03, Created)			
DR	01-MAY-1997 (TREMBLrel. 03, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Hexose oxidase.			
OS	Chondrus crispus (Carrageen).			
OC	Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineaceae;			
OX	Chondrus.			
NC	NCBI_TaxID-2769;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-97269074; PubMed-9111074;			
RA	Hansen O.C., Stougaard P.;			
RT	"Hexose oxidase from the red alga Chondrus crispus. Purification,			
RT	molecular cloning, and expression in Pichia pastoris.";			
RL	J. Biol. Chem. 272:11581-11587(1997).			
DR	EMBL: U089770; AAB0976.1;			
DR	InterPro: IPR001575; Oxid_FAD_bind.			
DR	Pfam: PF01565; FAD_binding_4; 1.			
SQ	SEQUENCE 546 AA; 61899 MW; 8DDDEA46D53C870B1 CRC64;			
Query Match	97.4%;	Score 75;	DB 10;	Length 546;
Best Local Similarity	93.8%;	Pred. No. 4.5e-05;		
Matches 15; Conservative	93.8%;	0; Mismatches 1;	Indels 0;	Gaps 0;
Qy	1 DLPMSRGVYASNLXF 16			
Db	219 DLPMSRGVYASNLXF 234			
RESULT 2	Q9D504	PRELIMINARY;	PRT;	1096 AA.
ID	Q9D504;			
AC	Q9D504;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			

```

DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DI 4921517D2IRIK protein.
GN 4921517D2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK014926; BAB29623.1; -
DR MGD: MGI:1914972; 4921517D2IRIK.
DR InterPro: IPR004000; Actin_like.
DR Pfam: PF00022; actin; 1.
DR SMART: SM00268; ACTIN; 1.
SQ SEQUENCE 1096 AA; 119041 MW; EB5E1F51D0A1C983 CRC64;

Query Match
Best Local Similarity 57.1%; Score 44; DB 11; Length 1096;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 LPMSPRGVIAS 12
DB 289 LPQSPRGVIAS 299

RESULT 3
OY09D5V1 PRELIMINARY; PRT; 1207 AA.
AC 09D5V1;
RT 01-JUN-2001 (TREMBlrel. 17, Created)
RT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
RT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DI 4921517D2IRIK protein.
GN 4921517D2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

```

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK014908; BAB29615.1; -
DR MGD: MGI:1914972; 4921517D2IRIK.
DR InterPro: IPR004000; Actin_like.
DR Pfam: PF00022; actin; 1.
DR SMART: SM00268; ACTIN; 1.
SQ SEQUENCE 1207 AA; 130803 MW; 2CA22CD754C0F5D8 CRC64;

Query Match
Best Local Similarity 57.1%; Score 44; DB 11; Length 1207;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 LPMSPRGVIAS 12
DB 400 LPQSPRGVIAS 410

RESULT 4
OY08U8M0 PRELIMINARY; PRT; 184 AA.
AC 08U8M0;
RT 01-JUN-2002 (TREMBlrel. 21, Created)
RT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
RT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DI 4921517D2IRIK protein.
GN 4921517D2IRIK.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Doan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RT Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Gurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaubin M., Iartchouk O., Epp A., Liu F.,
RA Flanagan C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
DR EMBL: AE009338; AA144871.1; -
DR EMBL: AE008277; AA89355.1; -
DR Complete proteome.
SQ SEQUENCE 184 AA; 19159 MW; 73D49565A448580B CRC64;

```



Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPMSPRGVIANLXF 16  
|||||  
Db 936 LPMSPRGVIANLNGF 950

## RESULT 8

ID 090235 PRELIMINARY; PRT; 229 AA.  
AC 090235;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Mx1 protein.  
OS Ambystoma mexicanum (Axolotl).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;  
OC Ambystoma.  
OX NCBI\_TaxID=8296;  
[1]

## SEQUENCE FROM N.A.

TISSUE-REGENERATING BLASTEMA;  
Koshiba K., Yamamoto H., Tamura K., Ide H., Kuroiwa A.;  
"Expression of Mx genes in regenerating and developing limbs of axolotl."  
Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC EMBL; DB2577; BAA11574.1; -  
DR HSSP; P14653; 1B72  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox.1.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR Prodom; PD000010; Homeobox.1.  
DR SMART; SM00389; HOX.1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
DR DNA-binding; Homeobox; Nuclear protein.  
KW SEQUENCE 229 AA; 25550 MW; 81BAD98CFE015 CRC64;

Query Match 51.9%; Score 40; DB 13; Length 229;  
Best Local Similarity 40.0%; Pred. No. 36;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 LPMSPRGVIANLXF 16  
:|||||:|  
Db 209 MPMPGGLAAMGY 223

## RESULT 9

ID 09HPM4 PRELIMINARY; PRT; 407 AA.  
AC 09HPM4;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE Vng1564h.  
GN VNG1564H.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
OX NCBI\_TaxID=64091;  
[1]

## SEQUENCE FROM N.A.

RX MEDLINE-20504483; PubMed-11016950;  
RA Ng W.-Y., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorson V., Sirogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonaki P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenberger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,

RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebbhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005068; AAC19843.1; -  
KW Complete proteome.  
SQ SEQUENCE 407 AA; 42994 MW; 2ED7DBB35C2C1380 CRC64;

Query Match 51.9%; Score 40; DB 17; Length 407;  
Best Local Similarity 46.2%; Pred. No. 65;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPMSPRGVIANSN 13  
|:|||||:|  
Db 223 DVALLPRGVANAS 235

## RESULT 10

ID 0974B5 PRELIMINARY; PRT; 682 AA.  
AC 0974B5;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative carbon-monoxide dehydrogenase large subunit.  
GN S70739.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
[1]

## SEQUENCE FROM N.A.

RP STRAIN-JCM 10545 / 7;  
RX PubMed-11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakayama H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7."  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000983; BAB65745.1; -  
DR InterPro; IPR000674; Aldoxan\_dh.C.  
DR InterPro; IPR001993; Mitoch\_carrier.  
DR Pfam; PF01315; Ald\_xan\_dh.C.1.  
DR Pfam; PF02738; Ald\_xan\_dh.C2.1.  
DR PROSITE; PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 682 AA; 74690 MW; 863B40E9B318AA6F CRC64;

Query Match 51.9%; Score 40; DB 17; Length 682;  
Best Local Similarity 70.0%; Pred. No. 11e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 PMSPGVIAN 12  
|||:|||||  
Db 172 PMEPRGVIAN 181

## RESULT 11

ID 09N9Z1 PRELIMINARY; PRT; 989 AA.  
AC 09N9Z1;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Endonuclease/reverse transcriptase.  
GN PILGERPOL OR POL.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea: Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOSON-PILGER NON-LTR RETROTRANSPOSON;  
 RA Tuschel T., Elbashir S., Lendackel W.;  
 RT "Drosophila melanogaster non-LTR retrotransposon pilger";  
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A278684; CAB99192.1;  
 DR FlyBase: FBgn0041726; pilger.pol.  
 DR InterPro: IPR005135; Exo\_endo\_phos.  
 DR InterPro: IPR000477; RVse.  
 DR Pfam: PF03372; Exo\_endo\_phos. 1.  
 DR Pfam: PF00078; rvc. 1.  
 KW RNA-directed DNA polymerase.  
 SQ SEQUENCE 989 AA; 110315 MW; 5957679C395D79A6 CRC64;  
 Query Match 51.9%; Score 40; DB 5; Length 989;  
 Best Local Similarity 60.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 PMSPRGVAS 12  
 DB 947 PLSPRGIVAA 956  
 RESULT 12  
 ID 092580 PRELIMINARY; PRT; 1193 AA.  
 AC 092580;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE KIAA0268 protein (Fragment).  
 GN KIAA0268.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE MARROW;  
 RA MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,  
 RA Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT Prediction of the coding sequences of unidentified human genes. VI.  
 RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain.";  
 EMBL: D87742; BAA13448.1;  
 DNA Res. 3:321-329(1996).  
 SQ SEQUENCE 1193 AA; 133704 MW; 70FB2542F9998038 CRC64;  
 NON\_TER 1  
 Query Match 51.9%; Score 40; DB 4; Length 1193;  
 Best Local Similarity 43.8%; Pred. No. 2e+02;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 OY 1 DLPMSPRGVASLXNF 16  
 DB 1113 DLPMPRGFLGHPAF 1128  
 RESULT 13  
 ID 0943R9 PRELIMINARY; PRT; 64 AA.  
 AC 0943R9;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE OSJNBa0083M16.17 protein.  
 GN OSJNBa0083M16.17.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Erihartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
 RT clone:OSJNBa0083M16.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003214; BAB64617.1;  
 SQ SEQUENCE 64 AA; 7195 MW; F0FD27343FC0C920 CRC64;  
 Query Match 50.6%; Score 39; DB 10; Length 64;  
 Best Local Similarity 87.5%; Pred. No. 15;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 LPMSPRGV 9  
 DB 5 LPSPRGV 12  
 RESULT 14  
 ID 0943K8 PRELIMINARY; PRT; 112 AA.  
 AC 0943K8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE P003D11.27 protein.  
 GN P003D11.27.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Erihartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P003D11.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003231; BAB67893.1;  
 SQ SEQUENCE 112 AA; 12233 MW; B750B7CA83A448D6 CRC64;  
 Query Match 50.6%; Score 39; DB 10; Length 112;  
 Best Local Similarity 87.5%; Pred. No. 26;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 LPMSPRGV 9  
 DB 5 LPSPRGV 12  
 RESULT 15  
 ID 091153 PRELIMINARY; PRT; 113 AA.  
 AC 091153;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Transcription factor (Fragment).  
 GN MSX-1.  
 OS Notoththalmus viridescens (Eastern newt) (Triturus viridescens).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;  
 OC Notoththalmus.  
 OX NCBI\_TaxID=8316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MID-BUD OF REGENERATING LIMB;  
 MEDLINE=95218226; PubMed=7703517;

RA Simon H.G., Nelson C., Goff D., Laufer E., Morgan B., Tabin C.;  
 RT "Differential expression of myogenic regulatory genes and Msx-1 during  
 RT dedifferentiation and redifferentiation of regenerating amphibian  
 RT limbs";  
 RL Dev. Dyn. 202;1-12(1995).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: X82837; CA58044.1; -.  
 DR HSSP: P22808; INK3.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00024; HOMEBOX.1.  
 DR ProDom: PD00010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX.1; 1.  
 DR PROSITE: PS50071; HOMEBOX.2; 1.  
 DR DNA-binding; Homeobox; Nuclear protein.  
 KW NON\_TER  
 FT  
 SQ SEQUENCE 113 AA; 12732 MW; 13E442H4858D982 CRC64;

Query Match 50.6%; Score 39; DB 13; Length 113;  
 Best Local Similarity 40.0%; Pred. No. 26;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

2 LPMSPRGVIAASLXNF 16  
 :||||| : : : : :  
 Db 93 MPMPGMYAAHVGY 107

## RESULT 16

Q8S724 PRELIMINARY; PRT; 150 AA.

AC Q8S724;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 16.6 kDa protein.  
 GN OSUNB0091N21.24.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_Taxid=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Buell C.R., Yuan O., Ouyang S., Liu J., Moffat K.S., Hill J.N.,  
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,  
 RA Tsitrin T., Riggs F., Hsiao J., Zisman V., Blunt S., Pal G.,  
 RA VanKen S.E., Utterback T.R., Feldblum T.V., Kalb E., Quackenbush J.,  
 RA Salzberg S.L., White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSUNB0091N21 genomic sequence.";  
 EMBL: AC091122; AAL82695.1; -.  
 Hypothetical Protein.  
 SEQUENCE 150 AA; 16595 MW; 9752C590F7DE4118 CRC64;

Query Match 50.6%; Score 39; DB 10; Length 150;  
 Best Local Similarity 87.5%; Pred. No. 35;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPMSPRGV 9  
 ||:|||||  
 Db 5 LPLSPRGV 12

## RESULT 17

Q91152 PRELIMINARY; PRT; 280 AA.

AC Q91152;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Msx-1 protein.

GN MSX-1.  
 OS Neothalpus viridescens (Eastern newt) (Triturus viridescens).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;  
 OC Neothalpus.  
 OX NCBI\_Taxid=8316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=REGENERATING BLASTEMA;  
 RX MEDLINE=95249617; PubMed=7732036;  
 RA Crews L., Gates P., Brown R., Joliot A., Foley C., Brookes J.P.,  
 RA Gann A.A.;  
 RT "Expression and activity of the newt Msx-1 gene in relation to limb  
 RT regeneration.";  
 RL Proc. R. Soc. Lond. B, Biol. Sci. 259:161-171(1995).  
 DR EMBL: X82395; CA57791.1; -.  
 DR HSSP: P14653; 1B72.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR ProDom: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS50071; HOMEBOX.2; 1.  
 KW  
 SQ SEQUENCE 280 AA; 30966 MW; B8BBD462003CA87B CRC64;

Query Match 50.6%; Score 39; DB 13; Length 280;  
 Best Local Similarity 40.0%; Pred. No. 67;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 LPMSPRGVIAASLXNF 16  
 :||||| : : : : :  
 Db 260 MPMPGMYAAHVGY 274

## RESULT 18

Q98116 PRELIMINARY; PRT; 442 AA.

AC Q98116;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein mlr2348.  
 GN MLR2348.  
 OS Rhizobium loti (Mesorhizobium loti).  
 ON Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bacteriota; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OX NCBI\_Taxid=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Saito S., Asamizu E., Kato T., Sasamoto S.,  
 RA Kishida Y., Kiyokawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP002999; BAB49500.1; -.  
 DR InterPro: IPR001145; Bac\_OmpA.  
 DR InterPro: IPR001035; OmpA; 1.  
 DR Pfam: PF00691; OmpA; 1.  
 DR PRINTS: PR01023; NAPLGMOTY.  
 DR ProDom: PD000930; Bac\_OmpA; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 442 AA; 47860 MW; 4853A307B1D71850 CRC64;

Query Match 50.6%; Score 39; DB 16; Length 442;  
 Best Local Similarity 60.0%; Pred. No. 11e+02;  
 Matches 9; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 1 DLPMSP--RGVIAASN 13



```

Db      205 DAPLSPMKGVIASD 219

RESULT 19
O9NMB8      PRELIMINARY;      PRT;      468 AA.
AC      O9NMB8;
DT      01-OCT-2000 (TREMblrel. 15, Created)
DT      01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT      01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE      Hypothetical 50.2 kDa protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=EMBRYO;
RA      Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA      Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA      WagaTsuna M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA      Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA      Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA      Yamamoto J., Wakematsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA      Niinomiya K., Iwayanagi T.;
RT      "NEDO human cDNA sequencing project.";
RL      Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK001005; BAA91465.1; -
DR      InterPro; IPR003597; Ig-cl.
DR      InterPro; IPR003600; Ig-like.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig_2.
DR      SMART; SM00407; IGcl; 1.
DR      SMART; SM00410; IGlike; 1.
DR      PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SO      SEQUENCE 468 AA; 50220 MW; 7C2FC6D5744FB101 CRC64;

Query Match      50.6%; Score 39; DB 4; Length 468;
Best Local Similarity 37.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY      1 DLPMSPRGVIASMLXF 16
Db      176 NLPLSPQGVTRAVEF 191

MULT 20
X59
O9BX59      PRELIMINARY;      PRT;      468 AA.
AC      O9BX59;
DT      01-JUN-2001 (TREMblrel. 17, Created)
DT      01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT      01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE      PLACE1007632 protein (Hypothetical 50.2 kDa protein).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=PLACENTA;
RA      Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA      Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA      WagaTsuna M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA      Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA      Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA      Yamamoto J., Wakematsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA      Niinomiya K., Iwayanagi T.;
RT      "NEDO human cDNA sequencing project.";
RL      Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.

```

```

RC      TISSUE=SKIN;
RA      Strausberg R.;
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK002056; BAB41077.1; -
DR      EMBL; BC015017; AAH15017.1; -
DR      InterPro; IPR003597; Ig.
DR      InterPro; IPR003597; Ig-cl.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig_2.
DR      SMART; SM00409; Ig; 1.
DR      SMART; SM00407; IGcl; 1.
DR      PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW      Hypothetical protein.
SO      SEQUENCE 468 AA; 50177 MW; 1826E2C8F3C39841 CRC64;

Query Match      50.6%; Score 39; DB 4; Length 468;
Best Local Similarity 37.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY      1 DLPMSPRGVIASMLXF 16
Db      176 NLPLSPQGVTRAVEF 191

Search completed: January 2, 2003, 12:07:20
Job time : 103 secs

```

